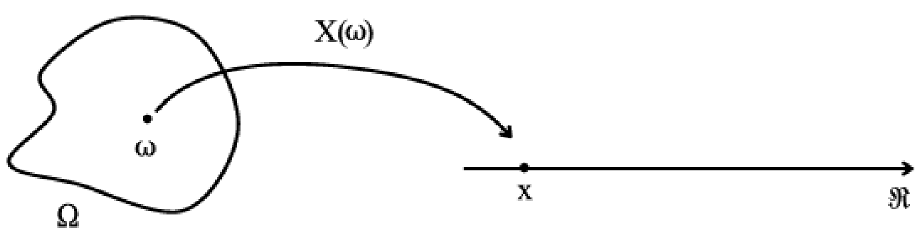
Review of Probability Theory

The focus of this course is on digital communication, which involves transmission of information, in its most general sense, from source to destination using digital technology. Engineering such a system requires modeling both the information and the transmission media. Interestingly, modeling both digital or analog information and many physical media requires a probabilistic setting. In this chapter and in the next one we will review the theory of probability, model random signals, and characterize their behavior as they traverse through deterministic systems disturbed by noise and interference. In order to develop practical models for random phenomena we start with carrying out a random experiment. We then introduce definitions, rules, and axioms for modeling within the context of the experiment. The outcome of a random experiment is denoted by ω. The sample space Ω is the set of all possible outcomes of a random experiment. Such outcomes could be an abstract description in words. A scientific experiment should indeed be repeatable where each outcome could naturally have an associated probability of occurrence. This is defined formally as the ratio of the number of times the outcome occurs to the total number of times the experiment is repeated.

Random Variables

A random variable is the assignment of a real number to each outcome of a random experiment.



Example:
Roll a dice. Outcomes ω ω ω ω ω ω
$\omega_i = i$ dots on the face of the dice.
X ω_i i

Distributions

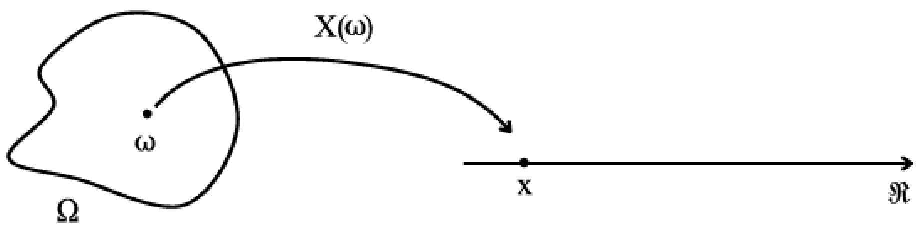
Probability assignments on intervals $a \quad X \quad b$

Cumulative distribution

The cumulative distribution function of a random variable X is a function $_X$ such that

Equation:

$$_X \; b \qquad X \quad b$$
$$\omega \quad \Omega \; X \; \omega \quad b$$



Continuous Random Variable

A random variable X is continuous if the cumulative distribution function can be written in an integral form, or

Equation:

$$_X \; b \qquad {}^{b} \quad _X \; x \quad x$$

and $_X \; x$ is the probability density function (pdf) (e.g., $_X \; x$ is differentiable and $_X \; x \quad \overline{}_{x} \quad _X \; x$)

Discrete Random Variable

A random variable X is discrete if it only takes at most countably many points (i.e., $_X$ is piecewise constant). The probability mass function (pmf) is defined as

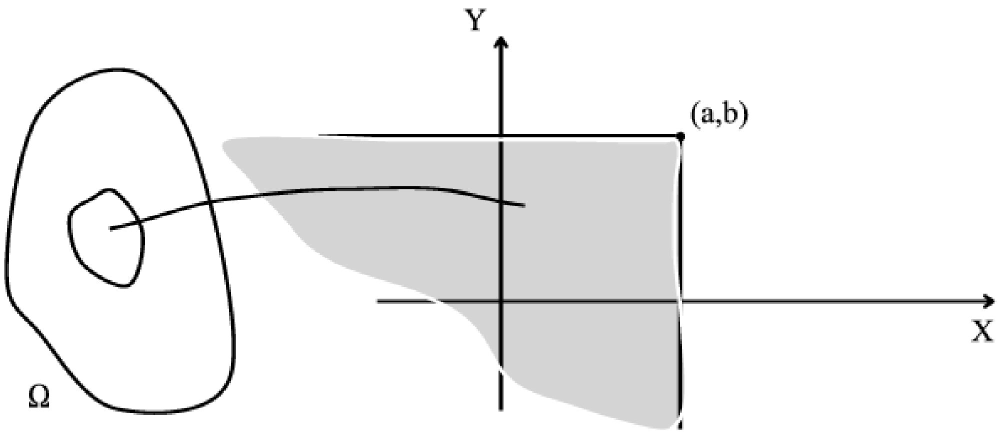
Equation:

$$_X \; x_k \qquad X \quad x_k$$
$$_X \; x_k \qquad _{x \; x \quad x_k \quad x \; x_k} \qquad _X \; x$$

Two random variables defined on an experiment have joint distribution

Equation:

$$_{X\ \ Y}\ \ a\ \ b \qquad\qquad X\quad a\ Y\quad b$$
$$\omega\quad \Omega\quad X\ \omega\quad a\qquad Y\ \omega\quad b$$



Joint pdf can be obtained if they are jointly continuous
Equation:

$$_{X\ \ Y}\ \ a\ \ b \qquad\qquad ^{b}\qquad ^{a}\qquad _{XY}\ \ x\ \ y\qquad x\qquad y$$

(e.g., $_{XY}\ \ x\ \ y \qquad \frac{_{X\ \ Y}\ x\ y}{x\ \ y}$)

Joint pmf if they are jointly discrete
Equation:

$$_{XY}\ \ x_k\ y_l \qquad\qquad X\quad x_k\ Y\quad y_l$$

Conditional density function
Equation:

$$f_{Y\ X}\ y\ x \qquad \frac{_{XY}\ \ x\ \ y}{_{X}\ \ x}$$

for all x with $_X \quad x$ otherwise conditional density is not defined for those values of x with $_X \quad x$

Two random variables are **independent** if
Equation:

$$_{XY} \; x \; y \qquad _X \; x \quad _Y \; y$$

for all x and y . For discrete random variables,
Equation:

$$_{XY} \; x_k \; y_l \qquad _X \; x_k \quad _Y \; y_l$$

for all k and l.

Moments

Statistical quantities to represent some of the characteristics of a random variable.
Equation:

$$g \; X \qquad E \; g \; X$$
$$g \; x \quad _X \; x \quad x$$
$$_k g \; x_k \quad _X \; x_k$$

- Mean
  Equation:

$$\mu_X \quad X$$

- Second moment
  Equation:

$$E \; X \qquad X$$

- Variance
  Equation:

$$X \quad \sigma \; X$$

$$X \quad \mu_X$$

$$X \quad \mu_X$$

- Characteristic function
  Equation:

  $$\Phi_X \; u \quad e^{iuX}$$

  for u , where i
- Correlation between two random variables
  Equation:

  $$R_{XY} \quad XY$$

  $$xy \quad {}_{X\,Y} \; x \; y \quad x \quad y$$

  $$_k \quad _l \, x_k y_l \quad {}_{X\,Y} \; x_k \; y_l$$

- Covariance
  Equation:

  $$C_{XY} \quad X \; Y$$

  $$X \quad \mu_X \quad Y \quad \mu_Y$$

  $$R_{XY} \quad \mu_X \mu_Y$$

- Correlation coefficient
  Equation:

  $$\rho_{XY} \quad \frac{X \; Y}{\sigma_X \sigma_Y}$$

Uncorrelated random variables

Two random variables X and Y are uncorrelated if ρ_{XY} .

Introduction to Stochastic Processes

Definitions, distributions, and stationarity

Stochastic Process

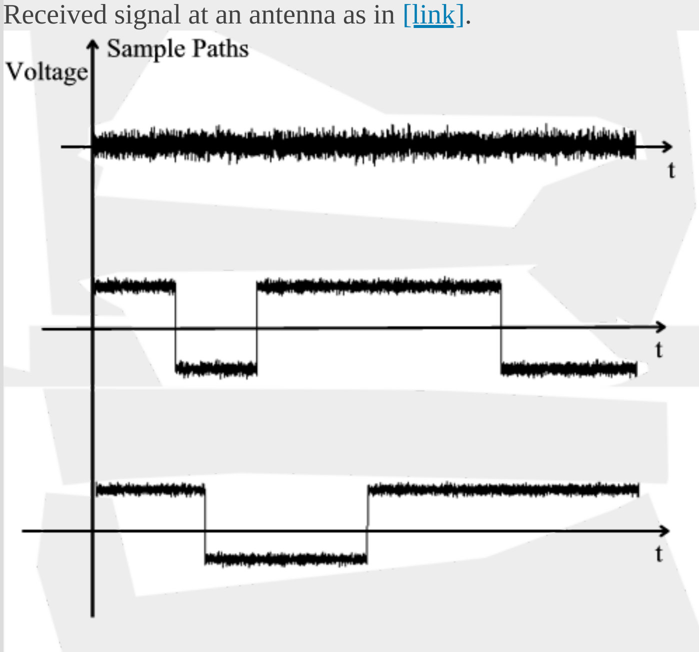
Given a sample space, a stochastic process is an indexed collection of random variables defined for each $\omega \in \Omega$.

Equation:

$$\forall t, t \in \quad : (X_t(\omega))$$

Example:

Received signal at an antenna as in [link].



For a given t, $X_t(\omega)$ is a random variable with a distribution

Equation:

First-order distribution

$$\begin{aligned} F_{X_t}(b) &= \Pr[X_t \leq b] \\ &= \Pr[\{\omega \in \Omega | X_t(\omega) \leq b\}] \end{aligned}$$

First-order stationary process

If $F_{X_t}(b)$ is not a function of time then X_t is called a first-order stationary process.

Equation:

Second-order distribution

$$F_{X_{t_1}, X_{t_2}}(b_1, b_2) = \Pr[X_{t_1} \leq b_1, X_{t_2} \leq b_2]$$

for all $t_1 \in$, $t_2 \in$, $b_1 \in$, $b_2 \in$

Equation:

Nth-order distribution

$$F_{X_{t_1}, X_{t_2}, \ldots, X_{t_N}}(b_1, b_2, \ldots, b_N) = \Pr[X_{t_1} \leq b_1, \ldots, X_{t_N} \leq b_N]$$

Nth-order stationary : A random process is stationary of order N if

Equation:

$$F_{X_{t_1}, X_{t_2}, \ldots, X_{t_N}}(b_1, b_2, \ldots, b_N) = F_{X_{t_1+T}, X_{t_2+T}, \ldots\ X_{t_N+T}}(b_1, b_2, \ldots, b_N)$$

Strictly stationary : A process is strictly stationary if it is Nth order stationary for all N.

Example:

$X_t = \cos(2\pi f_0 t + \Theta(\omega))$ where f_0 is the deterministic carrier frequency and $\Theta(\omega) : \Omega \to$ is a random variable defined over $[-\pi, \pi]$ and is assumed to be a uniform random variable; i.e.,

$$f_\Theta(\theta) = \begin{matrix} \frac{1}{2\pi} \text{ if } \theta \in [-\pi, \pi] \\ 0 \text{ otherwise} \end{matrix}$$

Equation:

$$\begin{aligned} F_{X_t}(b) &= \Pr[X_t \leq b] \\ &= \Pr[\cos(2\pi f_0 t + \Theta) \leq b] \end{aligned}$$

Equation:

$$F_{X_t}(b) = \Pr[-\pi \leq 2\pi f_0 t + \Theta \leq -\arccos(b)] + \Pr[\arccos(b) \leq 2\pi f_0 t + \Theta \leq \pi]$$

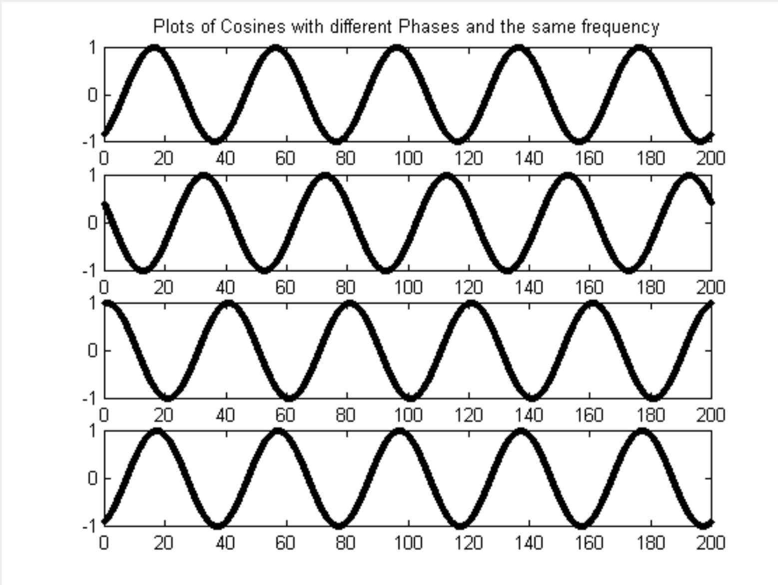
Equation:

$$\begin{aligned} F_{X_t}(b) &= \,_{(-\pi)-2\pi f_0 t}^{(-\arccos(b))-2\pi f_0 t} \frac{1}{2\pi}\, \mathrm{d}\,\theta + \,_{\arccos(b)-2\pi f_0 t}^{\pi-2\pi f_0 t} \frac{1}{2\pi}\, \mathrm{d}\,\theta \\ &= (2\pi - 2\arccos(b))\frac{1}{2\pi} \end{aligned}$$

Equation:

$$\begin{aligned} f_{X_t}(x) &= \frac{\mathrm{d}}{\mathrm{d}x}\ 1 - \frac{1}{\pi}\arccos(x) \\ &= \begin{matrix} \frac{1}{\pi\sqrt{1-x^2}} \text{ if } |x| \leq 1 \\ 0 \text{ otherwise} \end{matrix} \end{aligned}$$

This process is stationary of order 1.



The second order stationarity can be determined by first considering conditional densities and the joint density. Recall that

Equation:

$$X_t = \cos(2\pi f_0 t + \Theta)$$

Then the relevant step is to find

Equation:

$$\Pr[X_{t_2} \leq b_2 \mid X_{t_1} = x_1]$$

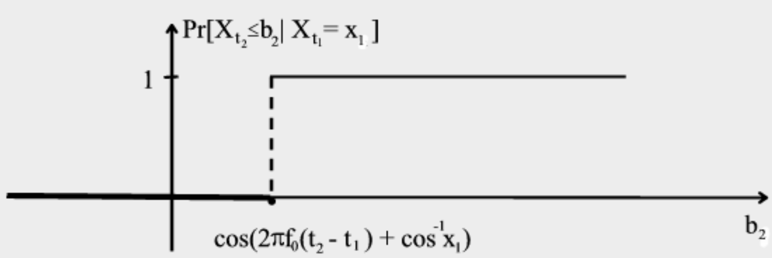
Note that

Equation:

$$(X_{t_1} = x_1 = \cos(2\pi f_0 t + \Theta)) \Rightarrow (\Theta = \arccos(x_1) - 2\pi f_0 t)$$

Equation:

$$\begin{aligned} X_{t_2} &= \cos(2\pi f_0 t_2 + \arccos(x_1) - 2\pi f_0 t_1) \\ &= \cos(2\pi f_0 (t_2 - t_1) + \arccos(x_1)) \end{aligned}$$

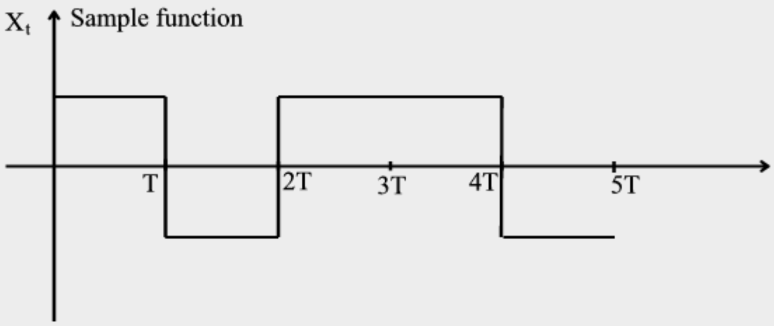


Equation:

$$F_{X_{t_2},X_{t_1}}(b_2,b_1) = \int_{-\infty}^{b_1} f_{X_{t_1}}(x_1)\,\Pr[X_{t_2} \le b_2 \mid X_{t_1} = x_1]\,\mathrm{d}\,x_1$$

Note that this is only a function of $t_2 - t_1$.

Example:
Every T seconds, a fair coin is tossed. If heads, then $X_t = 1$ for $nT \le t < (n+1)T$. If tails, then $X_t = -1$ for $nT \le t < (n+1)T$.



Equation:

$$p_{X_t}(x) = \begin{cases} \frac{1}{2} & \text{if } x = 1 \\ \frac{1}{2} & \text{if } x = -1 \end{cases}$$

for all $t \in$. X_t is stationary of order 1.
Second order probability mass function
Equation:

$$p_{X_{t_1}X_{t_2}}(x_1,x_2) = p_{X_{t_2}\ X_{t_1}}(x_2|x_1)\,p_{X_{t_1}}(x_1)$$

The conditional pmf
Equation:

$$p_{X_{t_2}\ X_{t_1}}(x_2|x_1) = \begin{cases} 0 & \text{if } x_2 \ne x_1 \\ 1 & \text{if } x_2 = x_1 \end{cases}$$

when $nT \le t_1 < (n+1)T$ and $nT \le t_2 < (n+1)T$ for some n.
Equation:

$$p_{X_{t_2}\ X_{t_1}}(x_2|x_1) = p_{X_t}\ (x_2)$$

for all x_1 and for all x_2 when $nT \le t_1 < (n+1)T$ and $mT \le t_2 < (m+1)T$ with $n \ne m$
Equation:

$$p_{X_{t_2}X_{t_1}}(x_2,x_1) = \begin{cases} 0 & \text{if } x_2 \ne x_1 \text{for } nT \le t_1, t_2 < (n+1)T \\ p_{X_{t_1}}(x_1) & \text{if } x_2 = x_1 \text{for } nT \le t_1, t_2 < (n+1)T \\ p_{X_{t_1}}(x_1)p_{X_{t_2}}(x_2) & \text{if } n \ne m \text{for } (nT \le t_1 < (n+1)T) \ \wedge\ (mT \le t_2 < (m+1)T) \end{cases}$$

Second-order description

Practical and incomplete statistics

Mean

The mean function of a random process X_t is defined as the expected value of X_t for all t's.

Equation:

$$\begin{aligned}\mu_{X_t} &= E[X_t] \\ &= \begin{cases}\int_{-\infty}^{\infty} x \, \mathrm{f}_{X_t}\ (x) \ \mathrm{d}\ x \ \ \text{if} \ \ \text{continuous} \\ \sum_{k}^{\infty}{}_{-\infty} x_k \, \mathrm{p}_{X_t}\ (x_k) \ \ \text{if} \ \ \text{discrete}\end{cases}\end{aligned}$$

Autocorrelation

The autocorrelation function of the random process X_t is defined as

Equation:

$$\begin{aligned}R_X(t_2, t_1) &= E \ \ X_{t_2}\overline{X_{t_1}} \\ &= \begin{cases}\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_2 \overline{x_1} \, \mathrm{f}_{X_{t_2}, X_{t_1}}\ (x_2, x_1) \ \mathrm{d}\ x_1 \ \mathrm{d}\ x_2 \ \ \text{if} \ \ \text{continuous} \\ \sum_{k}^{\infty}{}_{-\infty} \sum_{l}^{\infty}{}_{-\infty} x_l \overline{x_k} \, \mathrm{p}_{X_{t_2}, X_{t_1}}\ (x_l, x_k) \ \ \text{if} \ \ \text{discrete}\end{cases}\end{aligned}$$

Fact

If X_t is second-order stationary, then $R_X(t_2, t_1)$ only depends on $t_2 - t_1$.

Equation:

$$\begin{aligned}R_X(t_2, t_1) &= E \ \ X_{t_2}\overline{X_{t_1}} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_2 \overline{x_1} \, \mathrm{f}_{X_{t_2}, X_{t_1}}\ (x_2, x_1) \ \mathrm{d}\ x_2 \ \mathrm{d}\ x_1\end{aligned}$$

Equation:

$$
\begin{aligned}
R_X(t_2,t_1) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_2\overline{x_1}\ \mathrm{f}_{X_{t_2-t_1},X_0}\left(x_2,x_1\right)\mathrm{d}\,x_2\ \mathrm{d}\,x_1 \\
&= R_X(t_2-t_1,0)
\end{aligned}
$$

If $R_X(t_2,t_1)$ depends on t_2-t_1 only, then we will represent the autocorrelation with only one variable $\tau = t_2 - t_1$

Equation:

$$
\begin{aligned}
R_X(\tau) &= R_X(t_2-t_1) \\
&= R_X(t_2,t_1)
\end{aligned}
$$

Properties

1. $R_X(0) \geq 0$
2. $R_X(\tau) = \overline{R_X(-\tau)}$
3. $|R_X(\tau)| \leq R_X(0)$

Example:

$X_t = \cos(2\pi f_0 t + \Theta(\omega))$ and Θ is uniformly distributed between 0 and 2π. The mean function

Equation:

$$
\begin{aligned}
\mu_X(t) &= E[X_t] \\
&= E[\cos(2\pi f_0 t + \Theta)] \\
&= \int_0^{2\pi} \cos(2\pi f_0 t + \theta)\frac{1}{2\pi}\ \mathrm{d}\,\theta \\
&= 0
\end{aligned}
$$

The autocorrelation function

Equation:

$$
\begin{aligned}
R_X(t+\tau,t) &= E\ \ X_{t+\tau}\overline{X_t} \\
&= E[\cos(2\pi f_0\left(t+\tau\right)+\Theta)\cos(2\pi f_0 t+\Theta)] \\
&= 1/2E[\cos(2\pi f_0\tau)] + 1/2E[\cos(2\pi f_0\left(2t+\tau\right)+2\Theta)] \\
&= 1/2\cos(2\pi f_0\tau) + 1/2\int_0^{2\pi}\cos(2\pi f_0\left(2t+\tau\right)+2\theta)\frac{1}{2\pi}\ \mathrm{d}\,\theta \\
&= 1/2\cos(2\pi f_0\tau)
\end{aligned}
$$

Not a function of t since the second term in the right hand side of the equality in [link] is zero.

Example:
Toss a fair coin every T seconds. Since X_t is a discrete valued random process, the statistical characteristics can be captured by the pmf and the mean function is written as
Equation:

$$\begin{aligned} \mu_X(t) &= E[X_t] \\ &= 1/2 \times -1 + 1/2 \times 1 \\ &= 0 \end{aligned}$$

Equation:

$$\begin{aligned} R_X(t_2, t_1) &= \sum_{kk} \sum_{ll} x_k x_l \, \mathrm{p}_{X_{t_2}, X_{t_1}}(x_k, x_l) \\ &= 1 \times 1 \times 1/2 - 1 \times -1 \times 1/2 \\ &= 1 \end{aligned}$$

when $nT \leq t_1 < (n+1)T$ and $nT \leq t_2 < (n+1)T$
Equation:

$$\begin{aligned} R_X(t_2, t_1) &= 1 \times 1 \times 1/4 - 1 \times -1 \times 1/4 - 1 \times 1 \times 1/4 + 1 \times -1 \times 1/4 \\ &= 0 \end{aligned}$$

when $nT \leq t_1 < (n+1)T$ and $mT \leq t_2 < (m+1)T$ with $n \neq m$
Equation:

$$R_X(t_2, t_1) = \begin{cases} 1 \text{ if } & (nT \leq t_1 < (n+1)T) \wedge (nT \leq t_2 < (n+1)T) \\ 0 \text{ otherwise} \end{cases}$$

A function of t_1 and t_2.

Wide Sense Stationary
: A process is said to be wide sense stationary if μ_X is constant and $R_X(t_2, t_1)$ is only a function of $t_2 - t_1$.

Fact

If X_t is strictly stationary, then it is wide sense stationary. The converse is not necessarily true.

Autocovariance

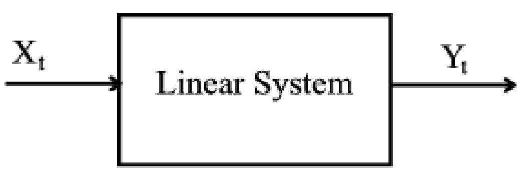
Autocovariance of a random process is defined as

Equation:

$$\begin{aligned} C_X(t_2,t_1) &= E\ (X_{t_2}-\mu_X(t_2))\overline{X_{t_1}-\mu_X(t_1)} \\ &= R_X(t_2,t_1)-\mu_X(t_2)\overline{\mu_X(t_1)} \end{aligned}$$

The variance of X_t is $\mathrm{Var}\,(X_t)=C_X(t,t)$

Two processes defined on one experiment ([link]).



Crosscorrelation

The crosscorrelation function of a pair of random processes is defined as

Equation:

$$\begin{aligned} R_{XY}(t_2,t_1) &= E\ X_{t_2}\overline{Y_{t_1}} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\,\mathrm{f}_{X_{t_2},Y_{t_1}}\ (x,y)\ \mathrm{d}\ x\ \mathrm{d}\ y \end{aligned}$$

Equation:

$$C_{XY}(t_2,t_1)=R_{XY}(t_2,t_1)-\mu_X(t_2)\overline{\mu_Y(t_1)}$$

Jointly Wide Sense Stationary

The random processes X_t and Y_t are said to be jointly wide sense stationary if $R_{XY}(t_2, t_1)$ is a function of $t_2 - t_1$ only and $\mu_X(t)$ and $\mu_Y(t)$ are constant.

Linear Filtering

Equation:

Integration

$$Z(\omega) = \int_{a}^{b} X_t(\omega)\ \mathrm{d}\ t$$

Equation:

Linear Processing

$$Y_t = \int_{-\infty}^{\infty} h(t,\tau) X_\tau\ \mathrm{d}\ \tau$$

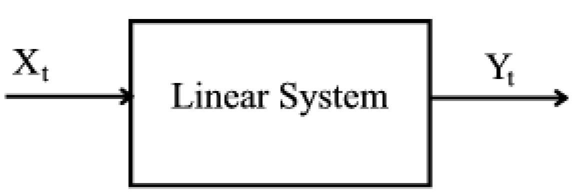
Equation:

Differentiation

$$X_t{}' = \frac{\mathrm{d}}{\mathrm{d}\ t}(X_t)$$

Properties

1. $Z = {}_{a}^{b} X_t(\omega)\ \mathrm{d}\ t = {}_{a}^{b} \mu_X(t)\ \mathrm{d}\ t$
2. $Z^2 = {}_{a}^{b} X_{t_2}\ \mathrm{d}\ t_2\ {}_{a}^{b} X_{t_1}\ \mathrm{d}\ t_1 = {}_{a}^{b}\ {}_{a}^{b} R_X(t_2, t_1)\ \mathrm{d}\ t_1\ \mathrm{d}\ t_2$



Equation:

$$
\begin{aligned}
\mu_Y(t) &= {}_{-\infty}^{\infty} h(t,\tau) X_\tau \,\mathrm{d}\,\tau \\
&= {}_{-\infty}^{\infty} h(t,\tau) \mu_X(\tau) \,\mathrm{d}\,\tau
\end{aligned}
$$

If X_t is wide sense stationary and the linear system is time invariant

Equation:

$$
\begin{aligned}
\mu_Y(t) &= {}_{-\infty}^{\infty} h(t-\tau) \mu_X \,\mathrm{d}\,\tau \\
&= \mu_X \; {}_{-\infty}^{\infty} h(t') \,\mathrm{d}\,t' \\
&= \mu_Y
\end{aligned}
$$

Equation:

$$
\begin{aligned}
R_{YX}(t_2,t_1) &= Y_{t_2} X_{t_1} \\
&= {}_{-\infty}^{\infty} h(t_2-\tau) X_\tau \,\mathrm{d}\,\tau X_{t_1} \\
&= {}_{-\infty}^{\infty} h(t_2-\tau) R_X(\tau-t_1) \,\mathrm{d}\,\tau
\end{aligned}
$$

Equation:

$$
\begin{aligned}
R_{YX}(t_2,t_1) &= {}_{-\infty}^{\infty} h(t_2-t_1-\tau') R_X(\tau') \,\mathrm{d}\,\tau' \\
&= h^* R_X(t_2-t_1)
\end{aligned}
$$

where $\tau' = \tau - t_1$.

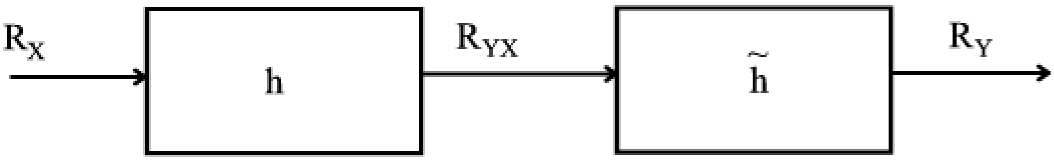
Equation:

$$
\begin{aligned}
R_Y(t_2,t_1) &= Y_{t_2} Y_{t_1} \\
&= Y_{t_2} \; {}_{-\infty}^{\infty} h(t_1,\tau) X_\tau \,\mathrm{d}\,\tau \\
&= {}_{-\infty}^{\infty} h(t_1,\tau) R_{YX}(t_2,\tau) \,\mathrm{d}\,\tau \\
&= {}_{-\infty}^{\infty} h(t_1-\tau) R_{YX}(t_2-\tau) \,\mathrm{d}\,\tau
\end{aligned}
$$

Equation:

$$
\begin{aligned}
R_Y(t_2, t_1) &= \int_{-\infty}^{\infty} h(\tau' - (t_2 - t_1)) R_{YX}(\tau')\, \mathrm{d}\,\tau' \\
&= R_Y(t_2 - t_1) \\
&= h^* R_{YX}(t_2, t_1)
\end{aligned}
$$

where $\tau' = t_2 - \tau$ and $h(\tau) = h(-\tau)$ for all $\tau \in$. Y_t is WSS if X_t is WSS and the linear system is time-invariant.



Example:
X_t is a wide sense stationary process with $\mu_X = 0$, and $R_X(\tau) = \frac{N_0}{2}\delta(\tau)$. Consider the random process going through a filter with impulse response $h(t) = e^{-(at)}u(t)$. The output process is denoted by Y_t. $\mu_Y(t) = 0$ for all t.
Equation:

$$
\begin{aligned}
R_Y(\tau) &= \frac{N_0}{2} \int_{-\infty}^{\infty} h(\alpha) h(\alpha - \tau)\, \mathrm{d}\,\alpha \\
&= \frac{N_0}{2} \frac{e^{-(a|\tau|)}}{2a}
\end{aligned}
$$

X_t is called a white process. Y_t is a Markov process.

Power Spectral Density

The power spectral density function of a wide sense stationary (WSS) process X_t is defined to be the Fourier transform of the autocorrelation function of X_t.
Equation:

$$S_X(f) = \int_{-\infty}^{\infty} R_X(\tau) e^{-(i2\pi f\tau)} \, \mathrm{d}\, \tau$$

if X_t is WSS with autocorrelation function $R_X(\tau)$.

Properties

1. $S_X(f) = S_X(-f)$ since R_X is even and real.
2. $\mathrm{Var}\,(X_t) = R_X(0) = {}_{-\infty}^{\infty} S_X(f) \, \mathrm{d}\, f$
3. $S_X(f)$ is real and nonnegative $S_X(f) \geq 0$ for all f.

If $Y_t = {}_{-\infty}^{\infty} h(t-\tau) X_\tau \, \mathrm{d}\, \tau$ then
Equation:

$$\begin{aligned} S_Y(f) &= (R_Y(\tau)) \\ &= \left(h^* h^* R_X(\tau)\right) \\ &= H(f) H(f) S_X(f) \\ &= (|H(f)|)^2 S_X(f) \end{aligned}$$

since $H(f) = {}_{-\infty}^{\infty} h(t) e^{-(i2\pi f t)} \, \mathrm{d}\, t = H(f)$

Example:
X_t is a white process and $h(t) = e^{-(at)} u(t)$.
Equation:

$$H(f) = \frac{1}{a + i2\pi f}$$

Equation:

$$S_Y(f) = \frac{\frac{N_0}{2}}{a^2 + 4\pi^2 f^2}$$

Gaussian Random Processes

Gaussian process

A process with mean $\mu_X(t)$ and covariance function $C_X(t_2, t_1)$ is said to be a Gaussian process if **any** $\boldsymbol{X} = (X_{t_1} X_{t_2} \ldots X_{t_N})^T$ formed by **any** sampling of the process is a Gaussian random vector, that is, **Equation:**

$$f_X(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{N}{2}} (\det \Sigma_X)^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_X)^T \Sigma_X{}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_X)}$$

for all $\boldsymbol{x} \in \ {}^n$ where

$$\boldsymbol{\mu}_X = \begin{matrix} \mu_X(t_1) \\ \vdots \\ \mu_X(t_N) \end{matrix}$$

and

$$\Sigma_X = \begin{matrix} C_X(t_1, t_1) & \ldots & C_X(t_1, t_N) \\ \vdots & \ddots & \\ C_X(t_N, t_1) & \ldots & C_X(t_N, t_N) \end{matrix}$$

. The complete statistical properties of X_t can be obtained from the second-order statistics.

Properties

1. If a Gaussian process is WSS, then it is strictly stationary.
2. If two Gaussian processes are uncorrelated, then they are also statistically independent.
3. Any linear processing of a Gaussian process results in a Gaussian process.

Example:
X and Y are Gaussian and zero mean and independent. $Z = X + Y$ is also Gaussian.
Equation:

$$\begin{aligned} \varphi_X(u) &= e^{iuX} \\ &= e^{-\frac{u^2}{2}\sigma_X^2} \end{aligned}$$

for all $u \in$
Equation:

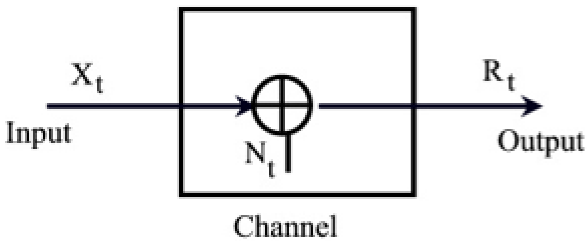
$$\begin{aligned} \varphi_Z(u) &= e^{iu(X+Y)} \\ &= e^{-\frac{u^2}{2}\sigma_X^2} e^{-\frac{u^2}{2}\sigma_Y^2} \\ &= e^{-\frac{u^2}{2}\sigma_X^2+\sigma_Y^2} \end{aligned}$$

therefore Z is also Gaussian.

Data Transmission and Reception

We will develop the idea of **data transmission** by first considering simple channels. In additional modules, we will consider more practical channels; **baseband** channels with **bandwidth** constraints and **passband** channels.

Simple additive white Gaussian channels



X_t carries data, N_t is a white Gaussian random process.

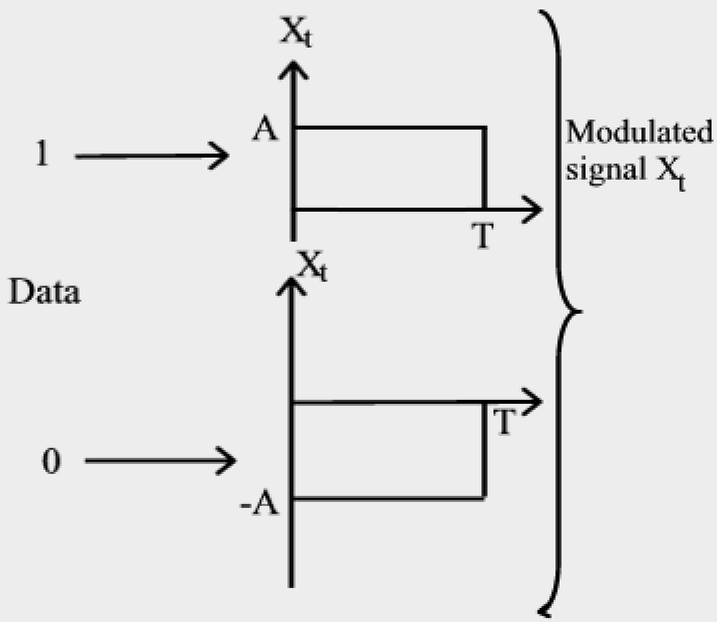
The concept of using different types of modulation for transmission of data is introduced in the module Signalling. The problem of demodulation and detection of signals is discussed in Demodulation and Detection.

Signalling

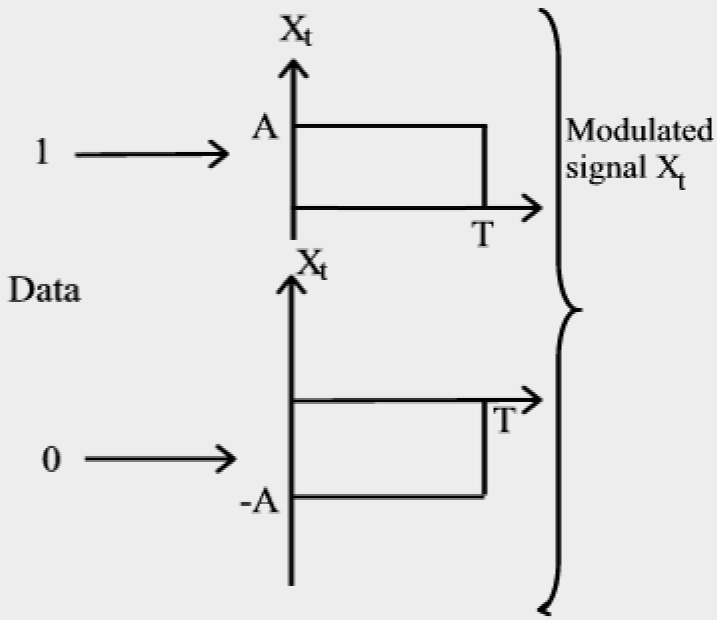
Example:

Data symbols are "1" or "0" and data rate is $\frac{1}{T}$ Hertz.

Pulse amplitude modulation (PAM)



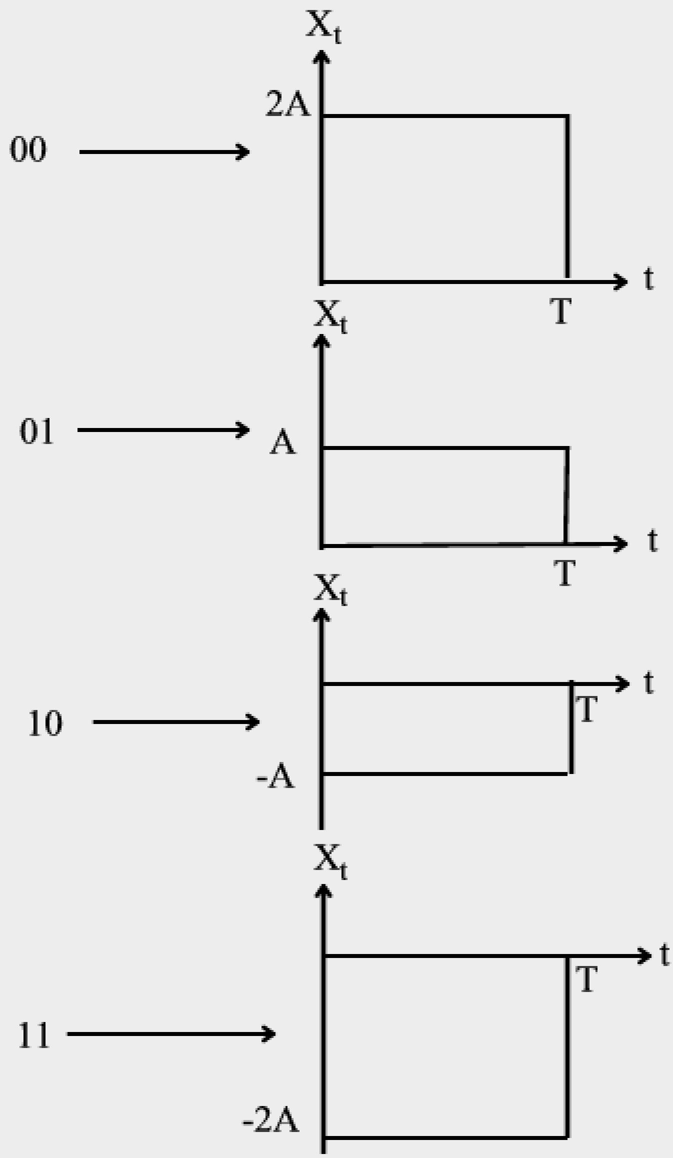
Pulse position modulation



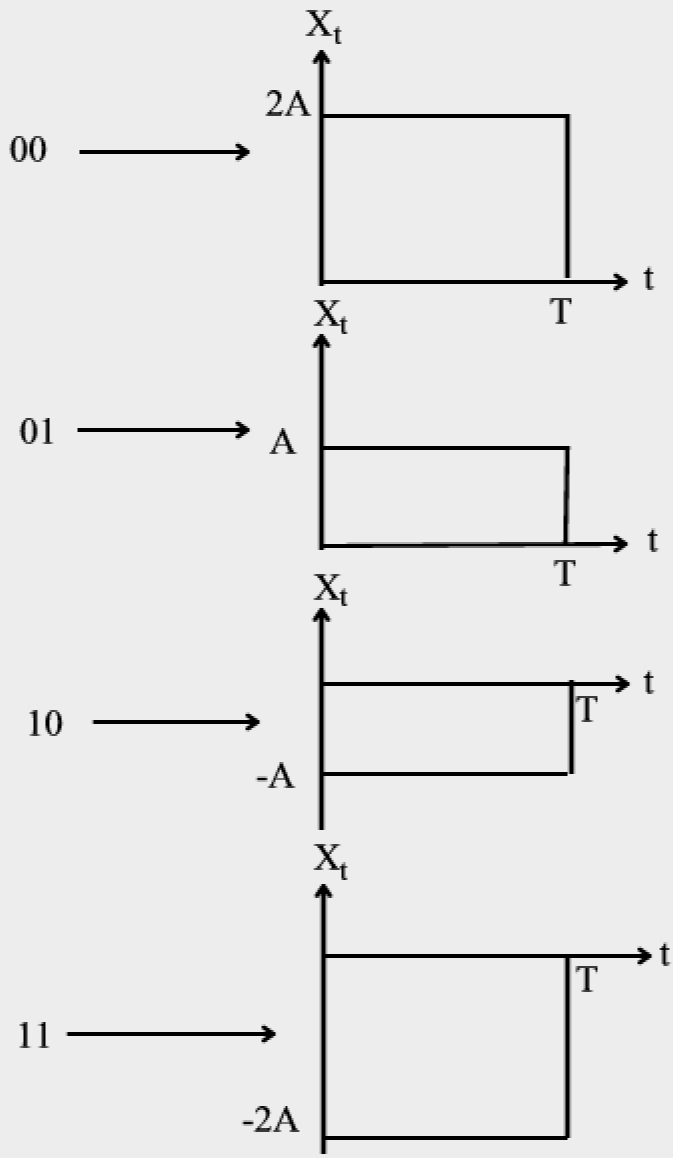
Example:

Example

Data symbols are "1" or "0" and the data rate is $\frac{2}{T}$ Hertz.



This strategy is an alternative to PAM with half the period, $\frac{T}{2}$.



Relevant measures are energy of modulated signals
Equation:

$$E_m = \forall m \in \{1, 2, \ldots, M\} : \int_0^T s_m{}^2(t)\,\mathrm{d}\,t$$

and how different they are in terms of inner products.

Equation:

$$\langle s_m, s_n \rangle = \int_0^T s_m(t) s_n(t)\, \mathrm{d}\, t$$

for $m \in \{1, 2, \ldots, M\}$ and $n \in \{1, 2, \ldots, M\}$.

antipodal

Signals $s_1(t)$ and $s_2(t)$ are antipodal if $\forall t, t \in [0, T] : (s_2(t) = -s_1(t))$

orthogonal

Signals $s_1(t)$, $s_2(t)$,…, $s_M(t)$ are orthogonal if $\langle s_m, s_n \rangle = 0$ for $m \neq n$.

biorthogonal

Signals $s_1(t)$, $s_2(t)$,…, $s_M(t)$ are biorthogonal if $s_1(t)$,…, $s_{\frac{M}{2}}(t)$ are orthogonal and $s_m(t) = -s_{\frac{M}{2}+m}(t)$ for some $m \in \left\{1, 2, \ldots, \frac{M}{2}\right\}$.

It is quite intuitive to expect that the smaller (the more negative) the inner products, $\langle s_m, s_n \rangle$ for all $m \neq n$, the better the signal set.

Simplex signals

Let $\{s_1(t), s_2(t), \ldots, s_M(t)\}$ be a set of orthogonal signals with equal energy. The signals $s_1(t)$,…, $s_M(t)$ are simplex signals if

Equation:

$$s_m(t) = s_m(t) - \frac{1}{M} \sum_{k=1}^{M} s_k(t)$$

If the energy of orthogonal signals is denoted by

Equation:

$$\forall m, m \in \{1, 2, \ldots, M\} : \quad E_s = \int_0^T {s_m}^2(t) \, \mathrm{d} \, t$$

then the energy of simplex signals
Equation:

$$E_{\tilde{s}} = \left(1 - \frac{1}{M}\right) E_s$$

and
Equation:

$$\forall m \neq n : \quad \langle s_m, s_n \rangle = \frac{-1}{M - 1} E_{\tilde{s}}$$

It is conjectured that among all possible M-ary signals with equal energy, the simplex signal set results in the smallest probability of error when used to transmit information through an additive white Gaussian noise channel.

The geometric representation of signals can provide a compact description of signals and can simplify performance analysis of communication systems using the signals.

Once signals have been modulated, the receiver must detect and demodulate the signals despite interference and noise and decide which of the set of possible transmitted signals was sent.

Geometric Representation of Modulation Signals

Geometric representation of signals can provide a compact characterization of signals and can simplify analysis of their performance as modulation signals.

Orthonormal bases are essential in geometry. Let $\{s_1(t), s_2(t), \ldots, s_M(t)\}$ be a set of signals.

Define $\psi_1(t) = \frac{s_1(t)}{\sqrt{E_1}}$ where $E_1 = \int_0^T {s_1}^2(t)\,\mathrm{d}\,t$.

Define $s_{21} = \langle s_2, \psi_1 \rangle = \int_0^T s_2(t)\psi_1(t)\,\mathrm{d}\,t$ and $\psi_2(t) = \frac{1}{\sqrt{E_2}}\left(s_2(t) - s_{21}\psi_1\right)$ where $E_2 = \int_0^T \left(s_2(t) - s_{21}\psi_1(t)\right)^2 \mathrm{d}\,t$

In general
Equation:

$$\psi_k(t) = \frac{1}{\sqrt{E_k}}\left(s_k(t) - \sum_{j=1}^{k-1} s_{\mathrm{kj}}\psi_j(t)\right)$$

where $E_k = \int_0^T \left(s_k(t) - \sum_{j=1}^{k-1} s_{\mathrm{kj}}\psi_j(t)\right)^2 \mathrm{d}\,t$.

The process continues until all of the M signals are exhausted. The results are N orthogonal signals with unit energy, $\{\psi_1(t), \psi_2(t), \ldots, \psi_N(t)\}$ where $N \leq M$. If the signals $\{s_1(t), \ldots, s_M(t)\}$ are linearly independent, then $N = M$.

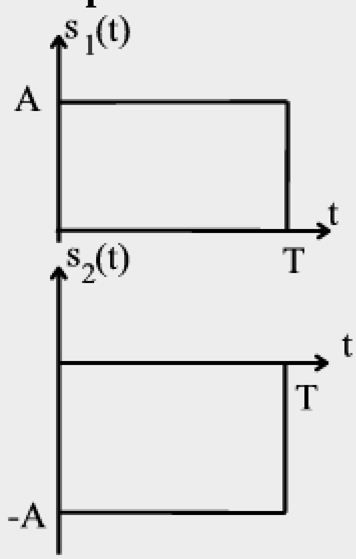
The M signals can be represented as
Equation:

$$s_m(t) = \sum_{n=1}^{N} s_{\mathrm{mn}}\psi_n(t)$$

with $m \in \{1, 2, \ldots, M\}$ where $s_{mn} = \langle s_m, \psi_n \rangle$ and $E_m = \sum_{n=1}^{N} {s_{mn}}^2$.

The signals can be represented by $s_m = \begin{pmatrix} s_{m1} \\ s_{m2} \\ \vdots \\ s_{mN} \end{pmatrix}$

Example:



Equation:

$$\psi_1(t) = \frac{s_1(t)}{\sqrt{A^2 T}}$$

Equation:

$$s_{11} = A\sqrt{T}$$

Equation:

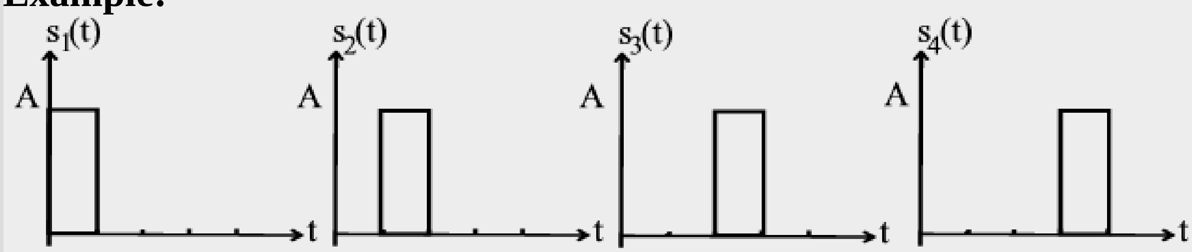
$$s_{21} = -A\sqrt{T}$$

Equation:

$$\begin{aligned}\psi_2(t) &= (s_2(t) - s_{21}\psi_1(t))\frac{1}{\sqrt{E_2}} \\ &= \left(-A + \frac{A\sqrt{T}}{\sqrt{T}}\right)\frac{1}{\sqrt{E_2}} \\ &= 0\end{aligned}$$

Dimension of the signal set is 1 with $E_1 = {s_{11}}^2$ and $E_2 = {s_{21}}^2$.

Example:



$\psi_m(t) = \frac{s_m(t)}{\sqrt{E_s}}$ where $E_s = \int_0^T {s_m}^2(t)\,\mathrm{d}t = \frac{A^2T}{4}$

$$s_1 = \begin{pmatrix}\sqrt{E_s} \\ 0 \\ 0 \\ 0\end{pmatrix},\ s_2 = \begin{pmatrix}0 \\ \sqrt{E_s} \\ 0 \\ 0\end{pmatrix},\ s_3 = \begin{pmatrix}0 \\ 0 \\ \sqrt{E_s} \\ 0\end{pmatrix},\ \text{and } s_4 = \begin{pmatrix}0 \\ 0 \\ 0 \\ \sqrt{E_s}\end{pmatrix}$$

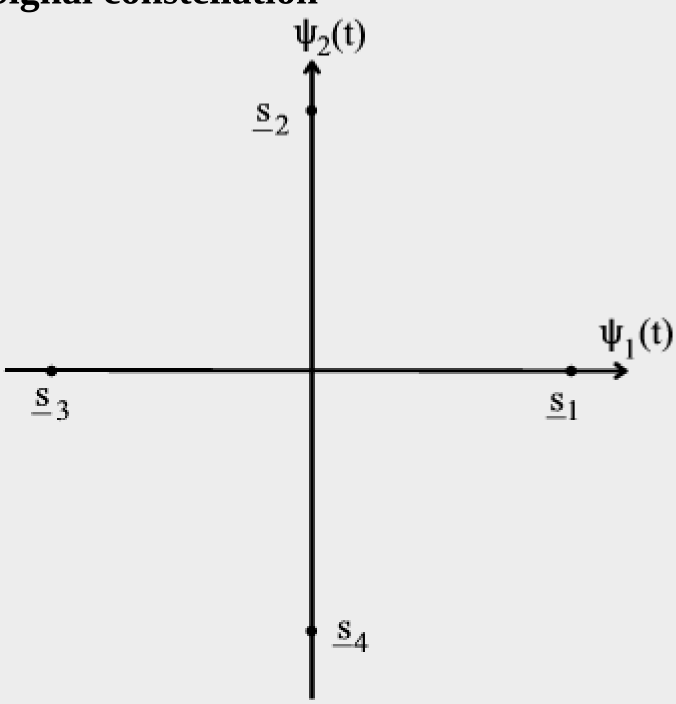
Equation:

$$\forall mn : \quad d_{\mathrm{mn}} = |s_m - s_n| = \sqrt{\sum_{j=1}^{N} (s_{\mathrm{mj}} - s_{\mathrm{nj}})^2} = \sqrt{2E_s}$$

is the Euclidean distance between signals.

Example:
Set of 4 equal energy biorthogonal signals. $s_1(t) = s(t)$, $s_2(t) = s^{\perp}(t)$, $s_3(t) = -s(t)$, $s_4(t) = -s^{\perp}(t)$.
The orthonormal basis $\psi_1(t) = \frac{s(t)}{\sqrt{E_s}}$, $\psi_2(t) = \frac{s^{\perp}(t)}{\sqrt{E_s}}$ where
$E_s = \int_0^T s_m{}^2(t)\, \mathrm{d}\, t$
$s_1 = \begin{pmatrix} \sqrt{E_s} \\ 0 \end{pmatrix}$, $s_2 = \begin{pmatrix} 0 \\ \sqrt{E_s} \end{pmatrix}$, $s_3 = \begin{pmatrix} -\sqrt{E_s} \\ 0 \end{pmatrix}$, $s_4 = \begin{pmatrix} 0 \\ -\sqrt{E_s} \end{pmatrix}$. The four signals can be geometrically represented using the 4-vector of projection coefficients s_1, s_2, s_3, and s_4 as a set of constellation points.

Signal constellation



Equation:

$$\begin{aligned} d_{21} &= |s_2 - s_1| \\ &= \sqrt{2E_s} \end{aligned}$$

Equation:

$$\begin{aligned} d_{12} &= d_{23} \\ &= d_{34} \\ &= d_{14} \end{aligned}$$

Equation:

$$\begin{aligned} d_{13} &= |s_1 - s_3| \\ &= 2\sqrt{E_s} \end{aligned}$$

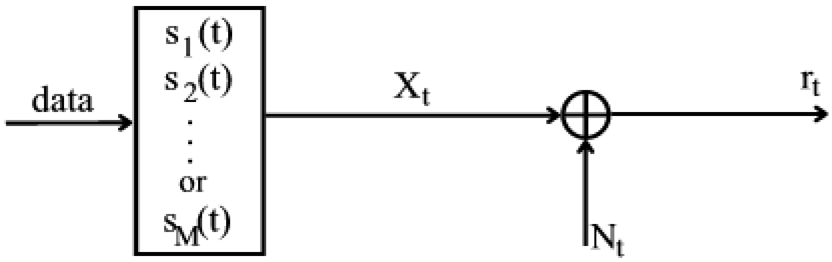
Equation:

$$d_{13} = d_{24}$$

Minimum distance $d_{\text{min}} = \sqrt{2E_s}$

Demodulation and Detection

Consider the problem where signal set, $\{s_1, s_2, ..., s_M\}$, for $t \in [0, T]$ is used to transmit $\log_2 M$ bits. The **modulated** signal X_t could be $\{s_1, s_2, ..., s_M\}$ during the interval $0 \le t \le T$.



$$r_t = X_t + N_t = s_m(t) + N_t \text{ for } 0 \le t \le T \text{ for some } m \in \{1, 2, ..., M\}.$$

Recall $s_m(t) = \sum_{n=1}^{N} s_{mn}\psi_n(t)$ for $m \in \{1, 2, ..., M\}$ the signals are decomposed into a set of orthonormal signals, perfectly.

Noise process can also be decomposed
Equation:

$$N_t = \sum_{n=1}^{N} \eta_n\psi_n(t) + \widetilde{N_t}$$

where $\eta_n = \int_0^T N_t\psi_n(t)\,\mathrm{d}\,t$ is the projection onto the n^{th} basis signal, $\widetilde{N_t}$ is the left over noise.

The problem of demodulation and detection is to observe r_t for $0 \le t \le T$ and decide which one of the M signals were transmitted. Demodulation is covered here. A discussion about detection can be found here.

Demodulation

Demodulation

Convert the continuous time received signal into a vector without loss of information (or performance).

Equation:

$$r_t = s_m(t) + N_t$$

Equation:

$$r_t = \sum_{n=1}^{N} s_{mn}\psi_n(t) + \sum_{n=1}^{N} \eta_n\psi_n(t) + \widetilde{N_t}$$

Equation:

$$r_t = \sum_{n=1}^{N} (s_{mn} + \eta_n)\psi_n(t) + \widetilde{N_t}$$

Equation:

$$r_t = \sum_{n=1}^{N} r_n\psi_n(t) + \widetilde{N_t}$$

The noise projection coefficients η_n's are zero mean, Gaussian random variables and are mutually independent if N_t is a white Gaussian process.

Equation:

$$\begin{aligned} \mu_\eta(n) &= E[\eta_n] \\ &= E\left[\int_0^T N_t\psi_n(t)\,\mathrm{d}\,t\right] \end{aligned}$$

Equation:

$$\begin{aligned}\mu_\eta(n) &= \int_0^T E[N_t]\psi_n(t)\,\mathrm{d}\,t \\ &= 0\end{aligned}$$

Equation:

$$\begin{aligned}E[\eta_k\eta_n] &= E\left[\int_0^T N_t\psi_k(t)\,\mathrm{d}\,t \int_0^T N_{t'}\psi_k(t')\,\mathrm{d}\,t'\right] \\ &= \int_0^T\int_0^T N_t N_{t'}\psi_k(t)\psi_n(t')\,\mathrm{d}\,t\,\mathrm{d}\,t'\end{aligned}$$

Equation:

$$E[\eta_k\eta_n] = \int_0^T\int_0^T R_N\left(t-t'\right)\psi_k(t)\psi_n\,\mathrm{d}\,t\,\mathrm{d}\,t'$$

Equation:

$$E[\eta_k\eta_n] = \frac{N_0}{2}\int_0^T\int_0^T \delta\left(t-t'\right)\psi_k(t)\psi_n(t')\,\mathrm{d}\,t\,\mathrm{d}\,t'$$

Equation:

$$\begin{aligned}E[\eta_k\eta_n] &= \frac{N_0}{2}\int_0^T \psi_k(t)\psi_n(t)\,\mathrm{d}\,t \\ &= \frac{N_0}{2}\delta_{kn} \\ &= \begin{cases}\frac{N_0}{2} \text{ if } k=n \\ 0 \text{ if } k\neq n\end{cases}\end{aligned}$$

η_k 's are uncorrelated and since they are Gaussian they are also independent. Therefore, $\eta_k \simeq \text{Gaussian}\left(0, \frac{N_0}{2}\right)$ and $R_\eta(k,n) = \frac{N_0}{2}\delta_{kn}$

The r_n's, the projection of the received signal r_t onto the orthonormal bases $\psi_n(t)$'s, are independent from the residual noise process $\widetilde{N_t}$.

The residual noise $\widetilde{N_t}$ is irrelevant to the decision process on r_t.

Recall $r_n = s_{mn} + \eta_n$, given $s_m(t)$ was transmitted. Therefore,

Equation:

$$\begin{aligned} \mu_r(n) &= E[s_{mn} + \eta_n] \\ &= s_{mn} \end{aligned}$$

Equation:

$$\begin{aligned} \mathrm{Var}\left(r_n\right) &= \mathrm{Var}\left(\eta_n\right) \\ &= \tfrac{N_0}{2} \end{aligned}$$

The correlation between r_n and $\widetilde{N_t}$

Equation:

$$E\left[\widetilde{N_t} r_n\right] = E \quad N_t - \sum_{k=1}^{N} \eta_k \psi_k(t) \quad s_{mn} + \eta_n$$

Equation:

$$E\left[\widetilde{N_t} r_n\right] = E \;\; N_t - \sum_{k=1}^{N} \eta_k \psi_k(t) \;\; s_{mn} + E[\eta_k \eta_n] - \sum_{k=1}^{N} E[\eta_k \eta_n] \psi_k(t)$$

Equation:

$$E\left[\widetilde{N_t} r_n\right] = E \;\; N_t \int_0^T N_{t'} \psi_n(t') \,\mathrm{d}\, t' \quad - \sum_{k=1}^{N} \frac{N_0}{2} \delta_{kn} \psi_k(t)$$

Equation:

$$E\left[\widetilde{N_t} r_n\right] = \int_0^T \frac{N_0}{2} \delta\left(t - t'\right) \psi_n\left(t'\right) \,\mathrm{d}\, t' - \frac{N_0}{2} \psi_n(t)$$

Equation:

$$
\begin{aligned}
E\left[\widetilde{N_t} r_n\right] &= \frac{N_0}{2}\psi_n(t) - \frac{N_0}{2}\psi_n(t) \\
&= 0
\end{aligned}
$$

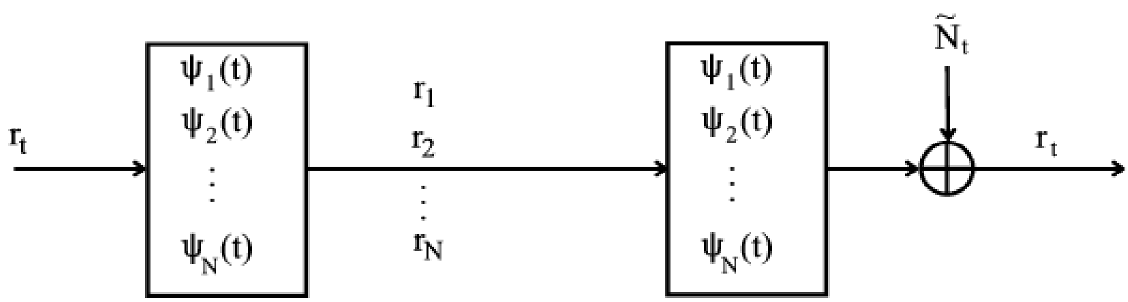
Since both $\widetilde{N_t}$ and r_n are Gaussian then $\widetilde{N_t}$ and r_n are also independent.

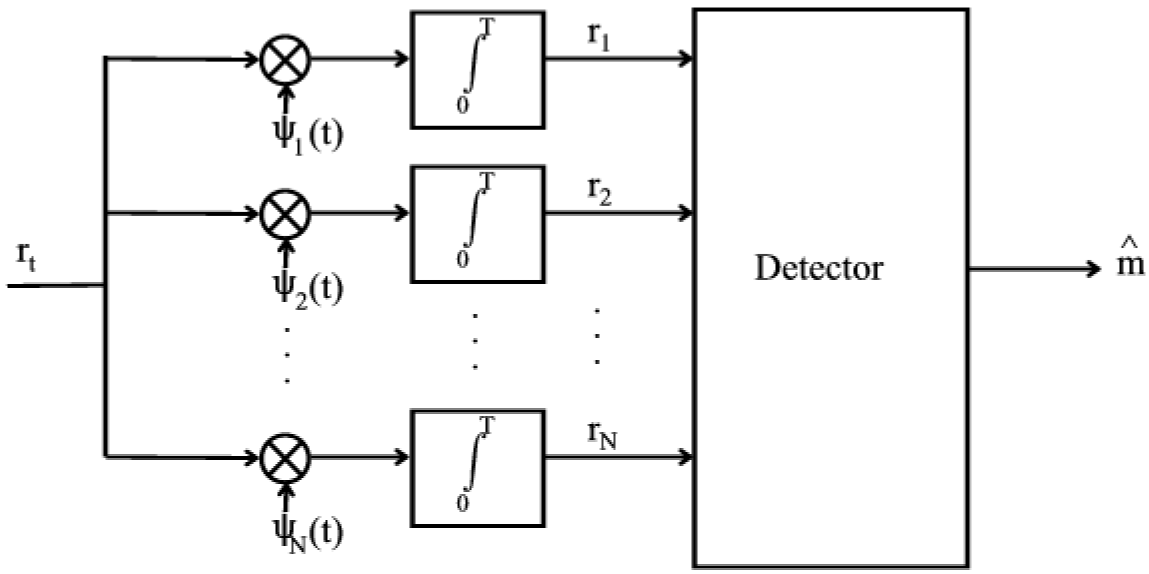
The conjecture is to ignore $\widetilde{N_t}$ and extract information from $\begin{matrix} r_1 \\ r_2 \\ \cdots \\ r_N \end{matrix}$.

Knowing the vector we can reconstruct the relevant part of random process r_t for $0 \le t \le T$

Equation:

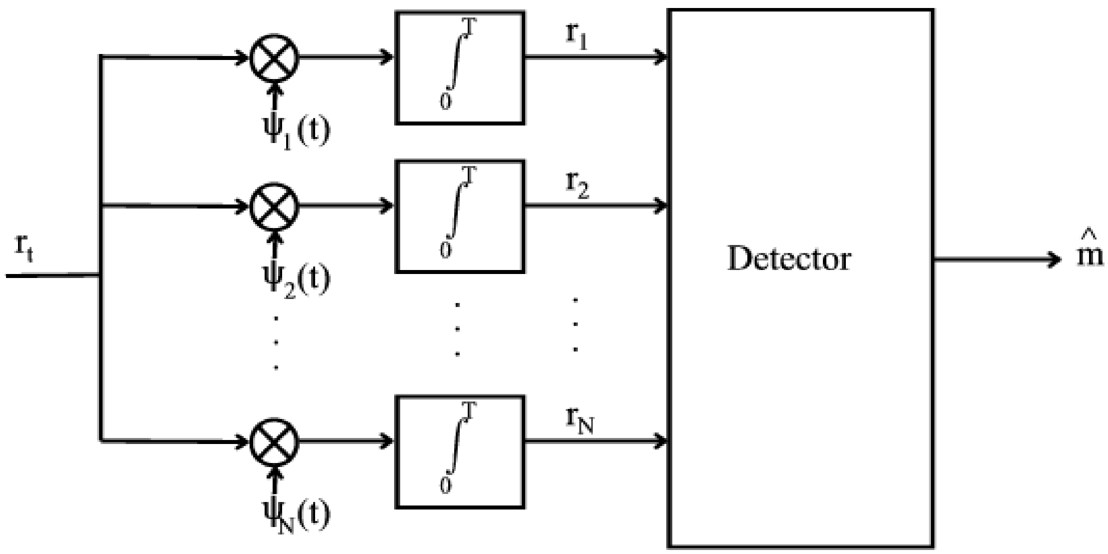
$$
\begin{aligned}
r_t &= s_m(t) + N_t \\
&= \sum_{n=1}^{N} r_n \psi_n(t) + \widetilde{N_t}
\end{aligned}
$$





Once the received signal has been converted to a vector, the correct transmitted signal must be detected based upon observations of the input vector. Detection is covered elsewhere.

Detection by Correlation
Demodulation and Detection



Detection

Decide which $s_m(t)$ from the set of $\{s_1(t), \ldots, s_m(t)\}$ signals was transmitted based on observing $= \begin{matrix} 1 \\ 2 \\ \vdots \\ N \end{matrix}$, the vector composed of [demodulated](#)received signal, that is, the vector of projection of the received signal onto the N bases.

Equation:

$$m = \arg\max_{1 \le m \le M} \Pr[s_m(t) \text{ was transmitted } \mid \quad \text{was observed}]$$

Note that

Equation:

$$\Pr[s_m \mid \quad] \quad \Pr[\ _m(t)\text{was transmitted} \mid \quad \text{was observed}] = \frac{f_{r|\ _m} \Pr[\ _m]}{f}$$

If $\Pr[\ _m \text{ was transmitted}] = \frac{1}{M}$, that is information symbols are equally likely to be transmitted, then

Equation:

$$\arg\max_{1 \le m \le M} \Pr[\ _m \mid \quad] = \arg\max_{1 \le m \le M} f_{\ |\ _m}$$

Since $r(t) = s_m(t) + N_t$ for $0 \le t \le T$ and for some $m = \{1, 2, \ldots, M\}$

then $= \ _m +$ where $= \begin{pmatrix} \ _1 \\ \ _2 \\ \vdots \\ \ _N \end{pmatrix}$ and $\ _n$'s are Gaussian and independent.

Equation:

$$\forall r_n, r_n \in \quad : \quad f_{\ |\ _m} = \frac{1}{2\pi \frac{N_0}{2}}^{\frac{N}{2}} e^{\frac{-\ _{n\ 1}^{N}\ _n - s_{m,n}\ ^2}{2\frac{N_0}{2}}}$$

Equation:

$$\begin{aligned} m &= \arg\max_{1 \le m \le M} f_{\ |\ _m} \\ &= \arg\max_{1 \le m \le M} \ln f_{\ |\ _m} \\ &= \arg\max_{1 \le m \le M} \quad - \frac{N}{2}\ln(\pi N_0) \quad - \frac{1}{N_0} \quad _{n\ 1}^{N} (\ _n - s_{m,n})^2 \\ &= \arg\min_{1 \le m \le M} \quad _{n\ 1}^{N} (\ _n - s_{m,n})^2 \end{aligned}$$

where $D(\ ,\ _m)$ is the l_2 distance between vectors and $\ _m$ defined as $D(\ ,\ _m) \quad _{n\ 1}^{N} (\ _n - s_{m,n})^2$

Equation:

$$\begin{aligned} m &= \arg\min_{1\leq m\leq M} D(\ ,\ _m) \\ &= \arg\min_{1\leq m\leq M} (\|\ \|)^2 - 2\langle(\ ,\ _m)\rangle + (\|\ _m\|)^2 \end{aligned}$$

where $\|\ \|$ is the l_2 norm of vector defined as $\|\ \|$ $\overline{{}^{N}_{n\ \ 1}(\ _n)^2}$

Equation:

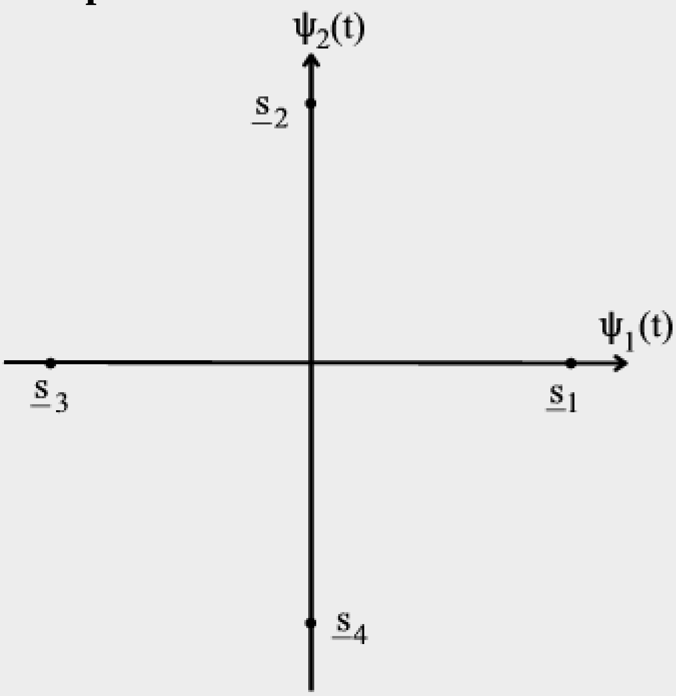
$$m = \arg\max_{1\leq m\leq M} 2\langle(\ ,\ _m)\rangle - (\|\ _m\|)^2$$

This type of receiver system is known as a **correlation** (or correlator-type) receiver. Examples of the use of such a system are found here. Another type of receiver involves linear, time-invariant filters and is known as a matched filter receiver. An analysis of the performance of a correlator-type receiver using antipodal and orthogonal binary signals can be found in Performance Analysis.

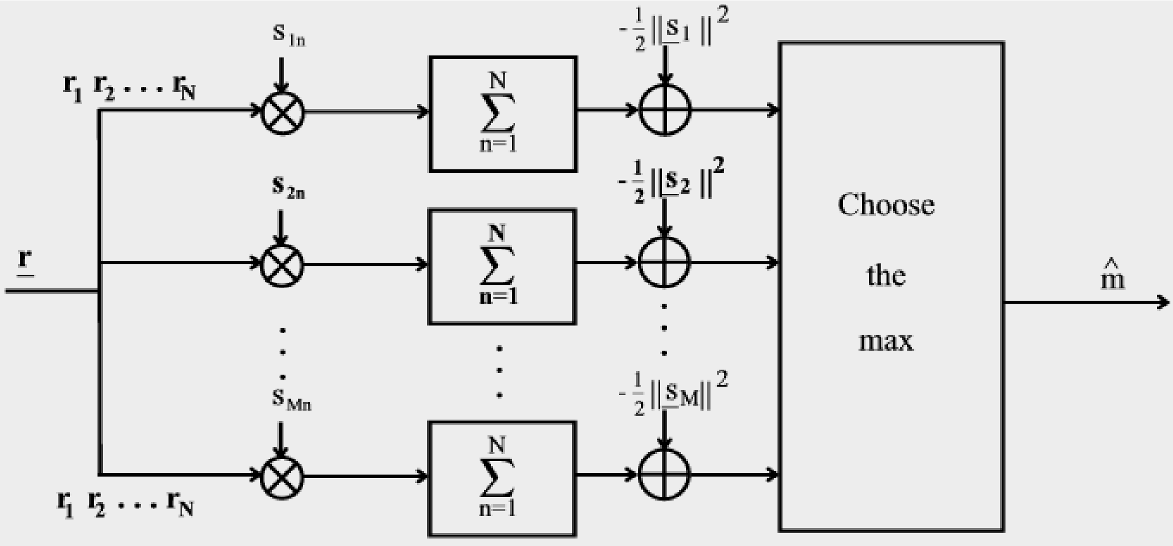
Examples of Correlation Detection

The implementation and theory of correlator-type receivers can be found in Detection.

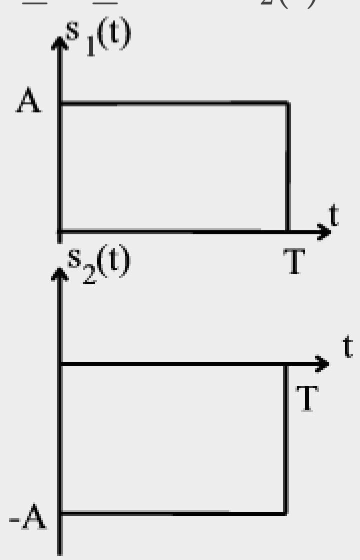
Example:



$\widehat{m} = 2$ since $D(\boldsymbol{r}, s_1) > D(\boldsymbol{r}, s_2)$ or $(\| s_1 \|)^2 = (\| s_2 \|)^2$ and $\langle \boldsymbol{r}, s_2 \rangle > \langle \boldsymbol{r}, s_1 \rangle$.



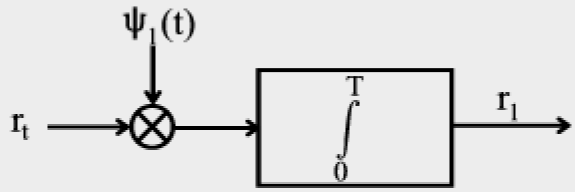
Example:
Data symbols "0" or "1" with equal probability. Modulator $s_1(t) = s(t)$ for $0 \leq t \leq T$ and $s_2(t) = -s(t)$ for $0 \leq t \leq T$.



$\psi_1(t) = \frac{s(t)}{\sqrt{A^2 T}}$, $s_{11} = A\sqrt{T}$, and $s_{21} = - \quad A\sqrt{T}$

Equation:

$$\forall m, m = \{1, 2\} : (r_t = s_m(t) + N_t)$$



Equation:

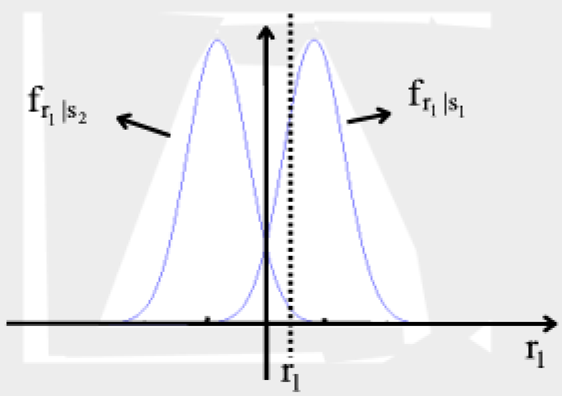
$$r_1 = A\sqrt{T} + \eta_1$$

or

Equation:

$$r_1 = -\quad A\sqrt{T}\quad + \eta_1$$

η_1 is Gaussian with zero mean and variance $\frac{N_0}{2}$.



$\widehat{m}$ =argmax $\quad A\sqrt{T}r_1, -\quad A\sqrt{T}r_1\quad$, since $A\sqrt{T} > 0$ and $\Pr[s_1] = \Pr[s_1]$ then the MAP decision rule decides.

$s_1(t)$ was transmitted if $r_1 \geq 0$

$s_2(t)$ was transmitted if $r_1 < 0$

An alternate demodulator:

Equation:

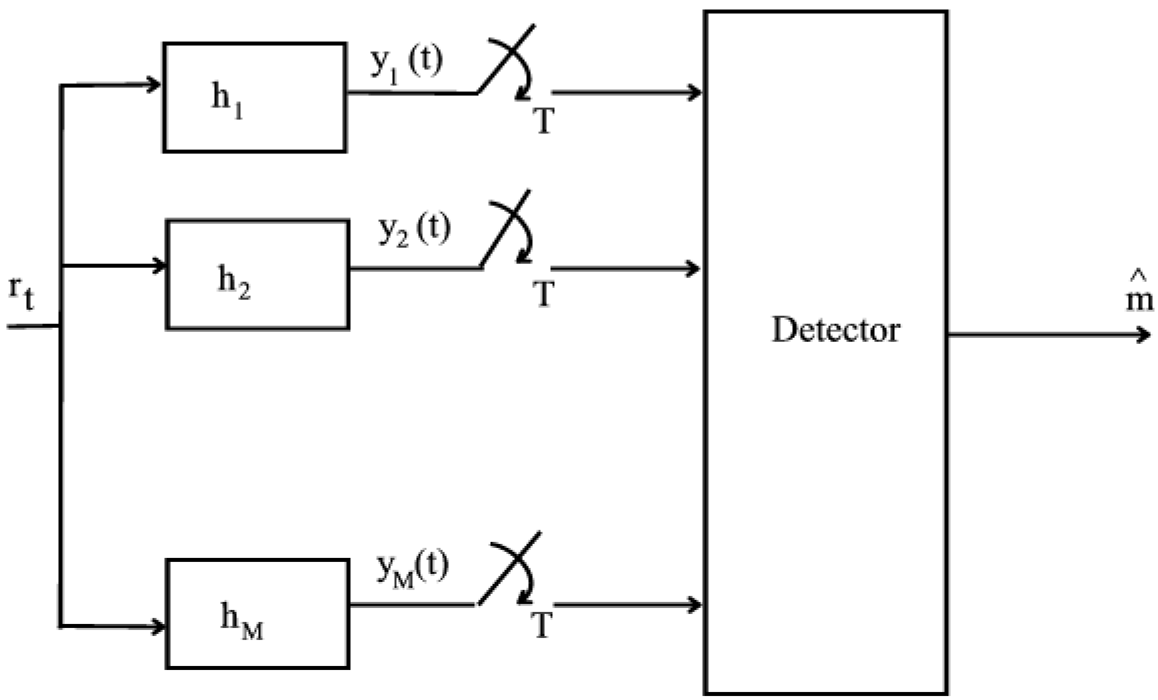
$$(r_t = s_m(t) + N_t) \Rightarrow (\boldsymbol{r} = s_m + \eta)$$

Matched Filters

Signal to Noise Ratio (SNR) at the output of the demodulator is a measure of the quality of the demodulator.
Equation:

$$\text{SNR} = \frac{\text{signal energy}}{\text{noise energy}}$$

In the correlator described earlier, $E_s = (|s_m|)^2$ and ${\sigma_{\eta_n}}^2 = \frac{N_0}{2}$. Is it possible to design a demodulator based on linear time-invariant filters with maximum signal-to-noise ratio?



If $s_m(t)$ is the transmitted signal, then the output of the k^{th} filter is given as
Equation:

$$\begin{aligned} y_k(t) &= \int_{-\infty}^{\infty} r_\tau h_k(t-\tau)\,\mathrm{d}\,\tau \\ &= \int_{-\infty}^{\infty} (s_m(\tau)+N_\tau)h_k(t-\tau)\,\mathrm{d}\,\tau \\ &= \int_{-\infty}^{\infty} s_m(\tau)h_k(t-\tau)\,\mathrm{d}\,\tau + \int_{-\infty}^{\infty} N_\tau h_k(t-\tau)\,\mathrm{d}\,\tau \end{aligned}$$

Sampling the output at time T yields

Equation:

$$y_k(T) = \int_{-\infty}^{\infty} s_m(\tau)h_k(T-\tau)\,\mathrm{d}\,\tau + \int_{-\infty}^{\infty} N_\tau h_k(T-\tau)\,\mathrm{d}\,\tau$$

The noise contribution:

Equation:

$$\nu_k = \int_{-\infty}^{\infty} N_\tau h_k(T-\tau)\,\mathrm{d}\,\tau$$

The expected value of the noise component is

Equation:

$$\begin{aligned} E[\nu_k] &= E\int_{-\infty}^{\infty} N_\tau h_k(T-\tau)\,\mathrm{d}\,\tau \\ &= 0 \end{aligned}$$

The variance of the noise component is the second moment since the mean is zero and is given as

Equation:

$$\begin{aligned} \sigma(\nu_k)^2 &= E\,{\nu_k}^2 \\ &= E\int_{-\infty}^{\infty} N_\tau h_k(T-\tau)\,\mathrm{d}\,\tau \int_{-\infty}^{\infty} N_{\tau'} h_k(T-\tau')\,\mathrm{d}\,\tau' \end{aligned}$$

Equation:

$$
\begin{aligned}
E\,\nu_k^{\,2} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{N_0}{2}\delta\ \tau - \tau \quad h_k(T-\tau)h_k(T-\tau\ )\ \mathrm{d}\,\tau\ \mathrm{d}\,\tau \\
&= \frac{N_0}{2}\int_{-\infty}^{\infty} (|h_k(T-\tau)|)^2\ \mathrm{d}\,\tau
\end{aligned}
$$

Signal Energy can be written as
Equation:

$$
\int_{-\infty}^{\infty} s_m(\tau)h_k(T-\tau)\ \mathrm{d}\,\tau \quad ^2
$$

and the signal-to-noise ratio (SNR) as
Equation:

$$
\mathrm{SNR} = \frac{\int_{-\infty}^{\infty} s_m(\tau)h_k(T-\tau)\ \mathrm{d}\,\tau \quad ^2}{\frac{N_0}{2}\int_{-\infty}^{\infty} (|h_k(T-\tau)|)^2\ \mathrm{d}\,\tau}
$$

The signal-to-noise ratio, can be maximized considering the well-known Cauchy-Schwarz Inequality
Equation:

$$
\int_{-\infty}^{\infty} g_1(x)g_2(x)\ \mathrm{d}\,x \quad ^2 \le \int_{-\infty}^{\infty} (|g_1(x)|)^2\ \mathrm{d}\,x \int_{-\infty}^{\infty} (|g_2(x)|)^2\ \mathrm{d}\,x
$$

with equality when $g_1(x) = \alpha g_2(x)$. Applying the inequality directly yields an upper bound on SNR
Equation:

$$
\frac{\int_{-\infty}^{\infty} s_m(\tau)h_k(T-\tau)\ \mathrm{d}\,\tau \quad ^2}{\frac{N_0}{2}\int_{-\infty}^{\infty} (|h_k(T-\tau)|)^2\ \mathrm{d}\,\tau} \le \frac{2}{N_0}\int_{-\infty}^{\infty} (|s_m(\tau)|)^2\ \mathrm{d}\,\tau
$$

with equality $\forall \tau : \quad h_k^{\mathrm{opt}}(T-\tau) = \alpha s_m(\tau)$. Therefore, the filter to examine signal m should be

Equation:

Matched Filter

$$\forall \tau : \quad h_m^{\mathrm{opt}}(\tau) = s_m(T-\tau)$$

The constant factor is not relevant when one considers the signal to noise ratio. The maximum SNR is unchanged when both the numerator and denominator are scaled.

Equation:

$$\frac{2}{N_0} \int_{-\infty}^{\infty} (|s_m(\tau)|)^2 \, d\tau = \frac{2E_s}{N_0}$$

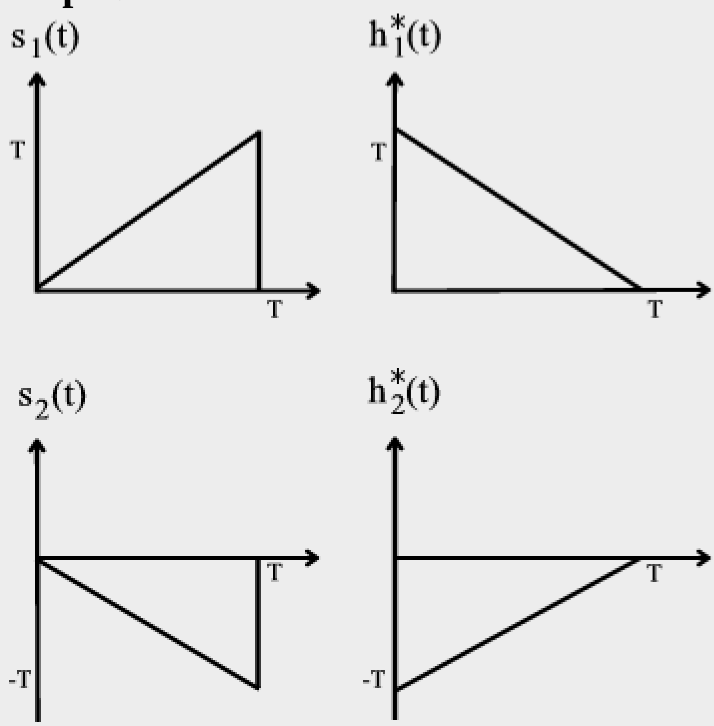
Examples involving matched filter receivers can be found here. An analysis in the frequency domain is contained in Matched Filters in the Frequency Domain.

Another type of receiver system is the correlation receiver. A performance analysis of both matched filters and correlator-type receivers can be found in Performance Analysis.

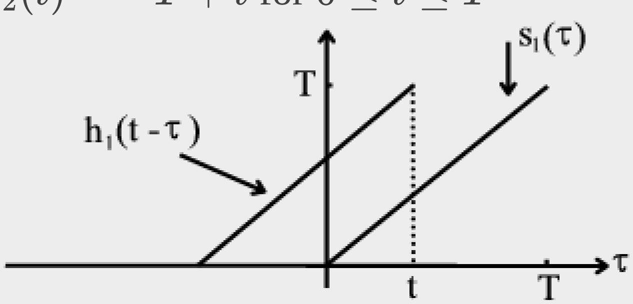
Examples with Matched Filters

The theory and rationale behind matched filter receivers can be found in Matched Filters.

Example:



$s_1(t) = t$ for $0 \le t \le T$
$s_2(t) = -t$ for $0 \le t \le T$
$h_1(t) = T - t$ for $0 \le t \le T$
$h_2(t) = -T + t$ for $0 \le t \le T$



Equation:

$$\forall t, 0 \leq t \leq 2T : \quad s_1(t) = \int_{-\infty}^{\infty} s_1(\tau) h_1(t-\tau)\, \mathrm{d}\,\tau$$

Equation:

$$\begin{aligned} s_1(t) &= \int_0^t \tau\,(T - t + \tau)\, \mathrm{d}\,\tau \\ &= \frac{1}{2}\,(T-t)\tau^2 \,\Big|_0^t + \frac{1}{3}\tau^3 \,\Big|_0^t \\ &= \frac{t^2}{2}\left(T - \frac{t}{3}\right) \end{aligned}$$

Equation:

$$s_1(T) = \frac{T^3}{3}$$

Compared to the correlator-type demodulation

Equation:

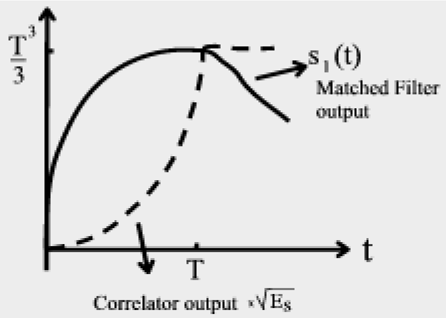
$$\psi_1(t) = \frac{s_1(t)}{\sqrt{E_s}}$$

Equation:

$$s_{11} = \int_0^T s_1(\tau)\psi_1(\tau)\, \mathrm{d}\,\tau$$

Equation:

$$\begin{aligned} \int_0^t s_1(\tau)\psi_1(\tau)\, \mathrm{d}\,\tau &= \frac{1}{\sqrt{E_s}} \int_0^t \tau\tau\, \mathrm{d}\,\tau \\ &= \frac{1}{\sqrt{E_s}}\,\frac{1}{3}t^3 \end{aligned}$$



Example:

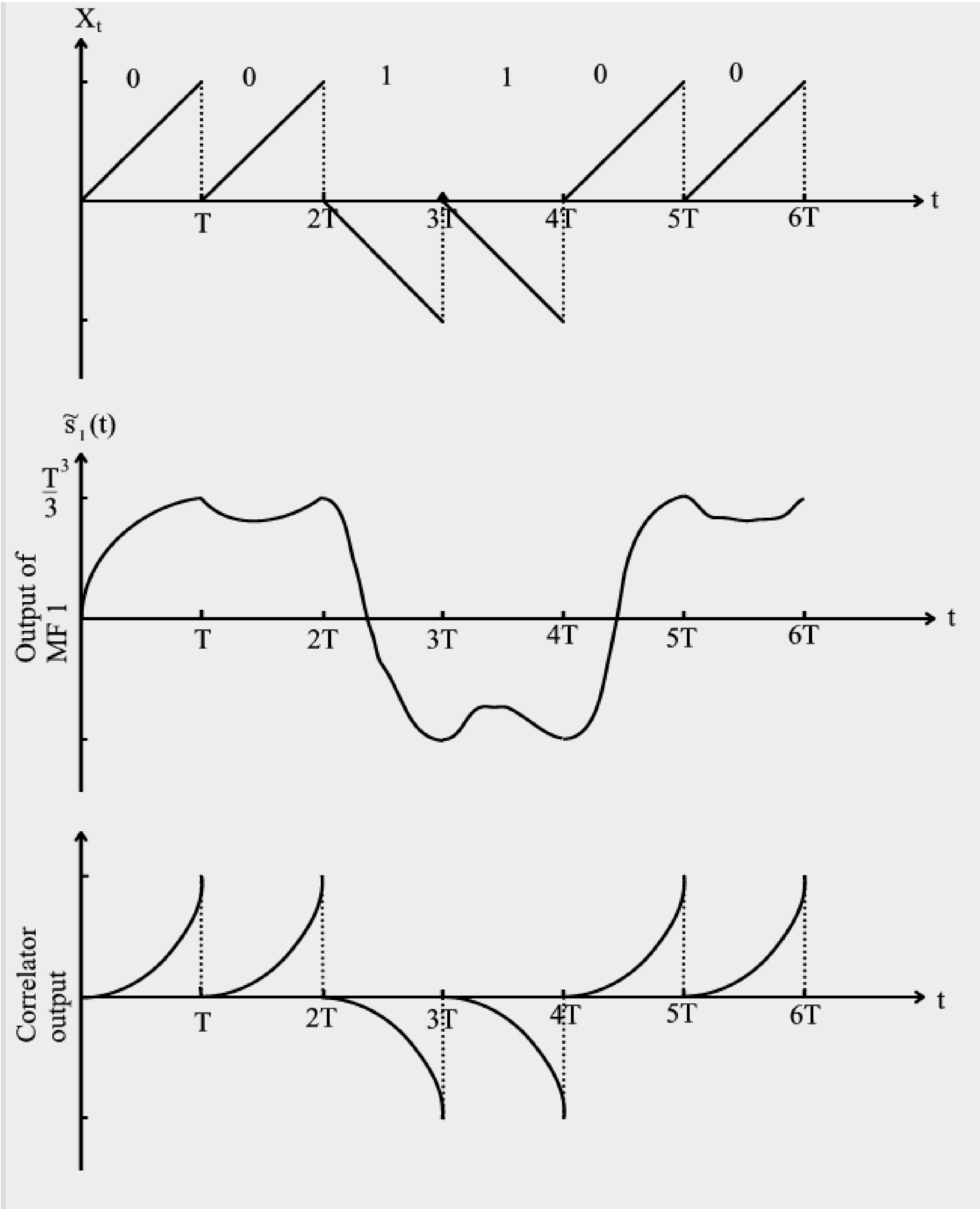
Assume binary data is transmitted at the rate of $\frac{1}{T}$ Hertz.

$0 \Rightarrow (b=1) \Rightarrow (s_1(t)=s(t))$ for $0 \leq t \leq T$

$1 \Rightarrow (b=-1) \Rightarrow (s_2(t)=-s(t))$ for $0 \leq t \leq T$

Equation:

$$X_t = \sum_{i=-P}^{P} b_i s(t-iT)$$



X_t
0
0
1
1
0
0
t
T
2T
3T
4T
5T
6T
$\tilde{s}_1(t)$
$\frac{T^3}{3}$
Output of MF 1
t
T
2T
3T
4T
5T
6T
Correlator output
t
T
2T
3T
4T
5T
6T

Matched Filters in the Frequency Domain

The time domain analysis and implementation of matched filters can be found in Matched Filters.

A frequency domain interpretation of matched filters is very useful
Equation:

$$\mathrm{SNR} = \frac{\left(\int_{-\infty}^{\infty} s_m(\tau) h_m(T-\tau)\,\mathrm{d}\,\tau\right)^2}{\frac{N_0}{2}\int_{-\infty}^{\infty} \left(|h_m(T-\tau)|\right)^2\,\mathrm{d}\,\tau}$$

For the m-th filter, h_m can be expressed as
Equation:

$$\begin{aligned} s_m(T) &= \int_{-\infty}^{\infty} s_m(\tau) h_m(T-\tau)\,\mathrm{d}\,\tau \\ &= \quad {}^{-1}(H_m(f) S_m(f)) \\ &= \int_{-\infty}^{\infty} H_m(f) S_m(f) e^{i2\pi fT}\,\mathrm{d}\,f \end{aligned}$$

where the second equality is because $\tilde{s}_m$ is the filter output with input S_m and filter H_m and we can now define $H_m(f) = H_m(f) e^{-(i2\pi fT)}$, then
Equation:

$$\tilde{s_m}(T) = \quad S_m(f), H_m(f)$$

The denominator
Equation:

$$\int_{-\infty}^{\infty} \left(|h_m(T-\tau)|\right)^2\,\mathrm{d}\,\tau = \int_{-\infty}^{\infty} \left(|h_m(\tau)|\right)^2\,\mathrm{d}\,\tau$$

Equation:

$$\begin{aligned} h_m * h_m(0) &= \int_{-\infty}^{\infty} (|H_m(f)|)^2 \, \mathrm{d} f \\ &= \langle H_m(f), H_m(f) \rangle \end{aligned}$$

Equation:

$$\begin{aligned} h_m * h_m(0) &= \int_{-\infty}^{\infty} H_m(f) e^{i2\pi fT} H_m(f) e^{-(i2\pi fT)} \, \mathrm{d} f \\ &= \quad H_m(f), H_m(f) \end{aligned}$$

Therefore,

Equation:

$$\mathrm{SNR} = \frac{\quad S_m(f), H_m(f) \quad ^2}{\frac{N_0}{2} \quad H_m(f), H_m(f) \quad} \leq \frac{2}{N_0} \langle (S_m(f), S_m(f)) \rangle$$

with equality when

Equation:

$$H_m(f) = \alpha S_m(f)$$

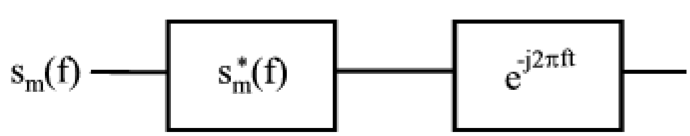
or

Equation:

Matched Filter in the frequency domain

$$H_m(f) = S_m(f) e^{-(i2\pi fT)}$$

Matched Filter



Equation:

$$\begin{aligned} s_m(t) &= {}^{-1}\; s_m(f) s_m(f) \\ &= \int_{-\infty}^{\infty} (|s_m(f)|)^2 e^{i2\pi ft}\, \mathrm{d}\, f \\ &= \int_{-\infty}^{\infty} (|s_m(f)|)^2 \cos(2\pi ft)\, \mathrm{d}\, f \end{aligned}$$

where ${}^{-1}$ is the inverse Fourier Transform operator.

Performance Analysis

In this section we will evaluate the probability of error of both correlator type receivers and matched filter receivers. We will only present the analysis for transmission of binary symbols. In the process we will demonstrate that both of these receivers have identical bit-error probabilities.

Antipodal Signals

$r_t = s_m(t) + N_t$ for $0 \leq t \leq T$ with $m = 1$ and $m = 2$ and $s_1(t) = -s_2(t)$

An analysis of the performance of correlation receivers with antipodal binary signals can be found here. A similar analysis for matched filter receivers can be found here.

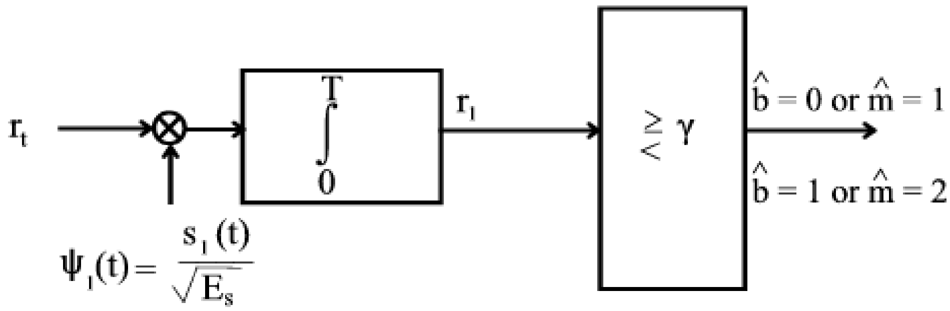
Orthogonal Signals

$r_t = s_m(t) + N_t$ for $0 \leq t \leq T$ with $m = 1$ and $m = 2$ and $\langle s_1, s_2 \rangle = 0$

An analysis of the performance of correlation receivers with orthogonal binary signals can be found here. A similar analysis for matched filter receivers can be found here.

It can be shown in general that correlation and matched filter receivers perform with the same symbol error probability if the detection criteria is the same for both receivers.

Performance Analysis of Antipodal Binary signals with Correlation



The bit-error probability for a correlation receiver with an antipodal signal set ([link]) can be found as follows:

Equation:

$$\begin{aligned} P_e &= \Pr\ m \neq m \\ &= \Pr\ \hat{b} \neq b \\ &= \pi_0 \Pr\ r_1 < \gamma|_{m=1}\ + \pi_1 \Pr\ r_1 \geq \gamma|_{m=2} \\ &= \pi_0\ {}_{-\infty}^{\gamma}\ \mathrm{f}_{r_1, s_1(t)}\ (r)\ \mathrm{d}\ r + \pi_1\ {}_{\gamma}^{\infty}\ \mathrm{f}_{r_1, s_2(t)}\ (r)\ \mathrm{d}\ r \end{aligned}$$

if $\pi_0 = \pi_1 = 1/2$, then the optimum threshold is $\gamma = 0$.

Equation:

$$\mathrm{f}_{r_1|s_1(t)}\ (r) = \qquad \overline{E_s}, \frac{N_0}{2}$$

Equation:

$$\mathrm{f}_{r_1|s_2(t)}\ (r) = \qquad - \ \overline{E_s}, \frac{N_0}{2}$$

If the two symbols are equally likely to be transmitted then $\pi_0 = \pi_1 = 1/2$ and if the threshold is set to zero, then

Equation:

$$P_e = 1/2 \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi \frac{N_0}{2}}} e^{-\frac{\left(r-\sqrt{E_s}\right)^2}{N_0}} \, \mathrm{d} r + 1/2 \int_{0}^{\infty} \frac{1}{\sqrt{2\pi \frac{N_0}{2}}} e^{-\frac{\left(r+\sqrt{E_s}\right)^2}{N_0}} \, \mathrm{d} r$$

Equation:

$$P_e = 1/2 \int_{-\infty}^{-\sqrt{\frac{2E_s}{N_0}}} \frac{1}{\sqrt{2\pi}} e^{\frac{-r'^2}{2}} \, \mathrm{d} r' + 1/2 \int_{\sqrt{\frac{2E_s}{N_0}}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-r''^2}{2}} \, \mathrm{d} r''$$

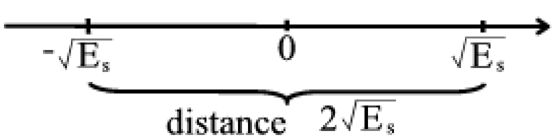
with $r' = \frac{r-\sqrt{E_s}}{\sqrt{\frac{N_0}{2}}}$ and $r'' = \frac{r+\sqrt{E_s}}{\sqrt{\frac{N_0}{2}}}$

Equation:

$$\begin{aligned} P_e &= \frac{1}{2} Q\left(\sqrt{\frac{2E_s}{N_0}}\right) + \frac{1}{2} Q\left(\sqrt{\frac{2E_s}{N_0}}\right) \\ &= Q\left(\sqrt{\frac{2E_s}{N_0}}\right) \end{aligned}$$

where $Q(b) = \int_b^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \, \mathrm{d} x$.

Note that



Equation:

$$P_e = Q\left(\frac{d_{12}}{\sqrt{2N_0}}\right)$$

where $d_{12} = 2\sqrt{E_s} = \left(\| \;_1 - \;_2 \|\right)^2$ is the Euclidean distance between the two constellation points ([link]).

This is exactly the same bit-error probability as for the matched filter case.

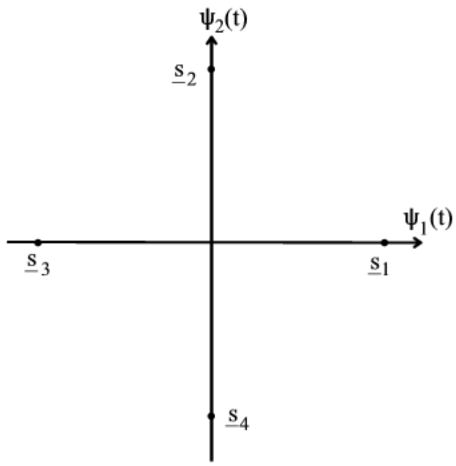
A similar bit-error analysis for matched filters can be found here. For the bit-error analysis for correlation receivers with an orthogonal signal set, refer here.

Performance Analysis of Binary Orthogonal Signals with Correlation

Orthogonal signals with equally likely bits, for , , , and .

Correlation (correlator-type) receiver

(see [link])



Decide was transmitted if .

Equation:

Equation:

Alternatively, if is transmitted we decide on the wrong signal if or or when .

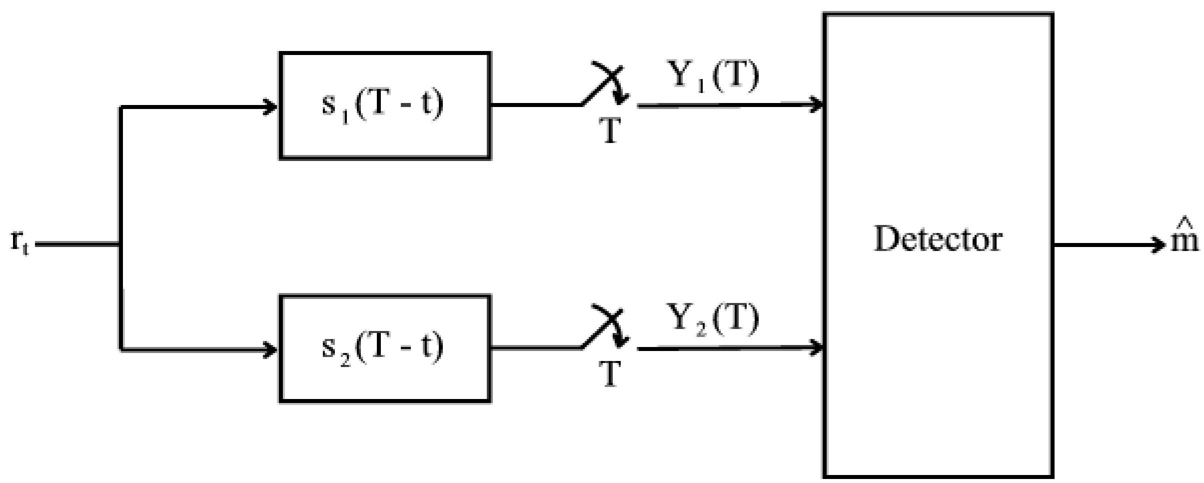
Equation:

Note that the distance between and is . The average bit error probability as we had for the antipodal case. Note also that the bit-error probability is the same as for the matched filter receiver.

Performance Analysis of Binary Antipodal Signals with Matched Filters

Matched Filter receiver

Recall $r_t = s_m(t) + N_t$ where $m = 1$ or $m = 2$ and $s_1(t) = -s_2(t)$ (see [link]).



Equation:

$$Y_1(T) = E_s + \nu_1$$

Equation:

$$Y_2(T) = -E_s + \nu_2$$

since $s_1(t) = -s_2(t)$ then ν_1 is $\mathcal{N}\left(0, \frac{N_0}{2} E_s\right)$. Furthermore $\nu_2 = -\nu_1$. Given ν_1 then ν_2 is deterministic and equals $-\nu_1$. Then $Y_2(T) = -Y_1(T)$ if $s_1(t)$ is transmitted.

If $s_2(T)$ is transmitted
Equation:

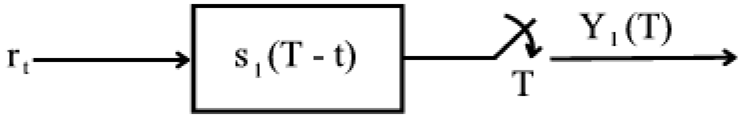
$$Y_1(T) = -E_s + \nu_1$$

Equation:

$$Y_2(T) = E_s + \nu_2$$

ν_1 is $\mathcal{N}\left(0, \frac{N_0}{2} E_s\right)$ and $\nu_2 = -\nu_1$.

The receiver can be simplified to (see [link])



If $s_1(t)$ is transmitted $Y_1(T) = E_s + \nu_1$.

If $s_2(t)$ is transmitted $Y_1(T) = -E_s + \nu_1$.

Equation:

$$\begin{aligned} P_e &= 1/2 \Pr[Y_1(T) < 0 \mid s_1(t)] + 1/2 \Pr[Y_1(T) \geq 0 \mid s_2(t)] \\ &= 1/2 \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi \frac{N_0}{2} E_s}} e^{\frac{-(|y - E_s|)^2}{N_0 E_s}} \, \mathrm{d}\, y + 1/2 \int_{0}^{\infty} \frac{1}{\sqrt{2\pi \frac{N_0}{2} E_s}} e^{\frac{-(|y + E_s|)^2}{N_0 E_s}} \, \mathrm{d}\, y \\ &= Q\left(\frac{E_s}{\sqrt{\frac{N_0}{2} E_s}}\right) \\ &= Q\left(\sqrt{\frac{2E_s}{N_0}}\right) \end{aligned}$$

This is the exact bit-error rate of a correlation receiver. For a bit-error analysis for orthogonal signals using a matched filter receiver, refer here.

Performance Analysis of Orthogonal Binary Signals with Matched Filters

Equation:

$$r_t \Rightarrow \quad = \begin{array}{c} Y_1(T) \\ Y_2(T) \end{array}$$

If $s_1(t)$ is transmitted

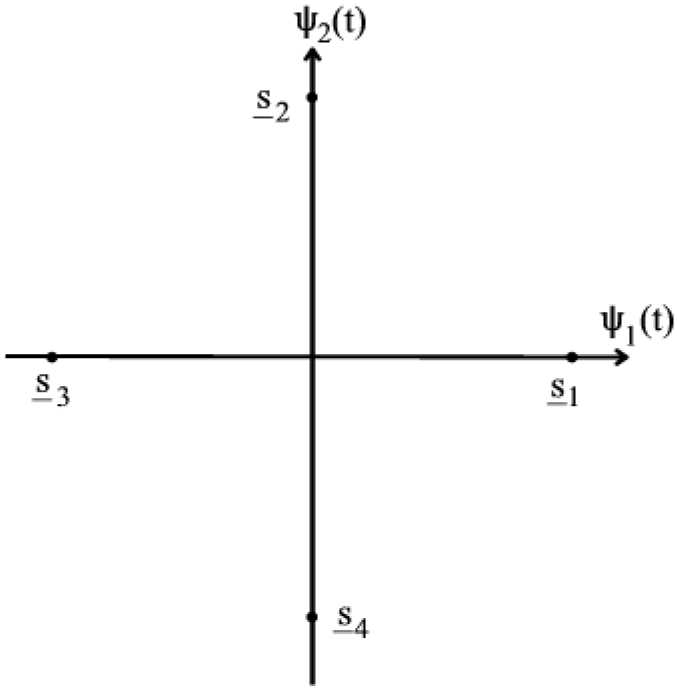
Equation:

$$\begin{aligned} Y_1(T) &= \int_{-\infty}^{\infty} s_1(\tau) h_1^{\mathrm{opt}}(T-\tau)\,\mathrm{d}\,\tau + \nu_1(T) \\ &= \int_{-\infty}^{\infty} s_1(\tau) s_1^*(\tau)\,\mathrm{d}\,\tau + \nu_1(T) \\ &= E_s + \nu_1(T) \end{aligned}$$

Equation:

$$\begin{aligned} Y_2(T) &= \int_{-\infty}^{\infty} s_1(\tau) s_2^*(\tau)\,\mathrm{d}\,\tau + \nu_2(T) \\ &= \nu_2(T) \end{aligned}$$

If $s_2(t)$ is transmitted, $Y_1(T) = \nu_1(T)$ and $Y_2(T) = E_s + \nu_2(T)$.



Equation:

H0

$$= \begin{pmatrix} E_s \\ 0 \end{pmatrix} + \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix}$$

Equation:

H1

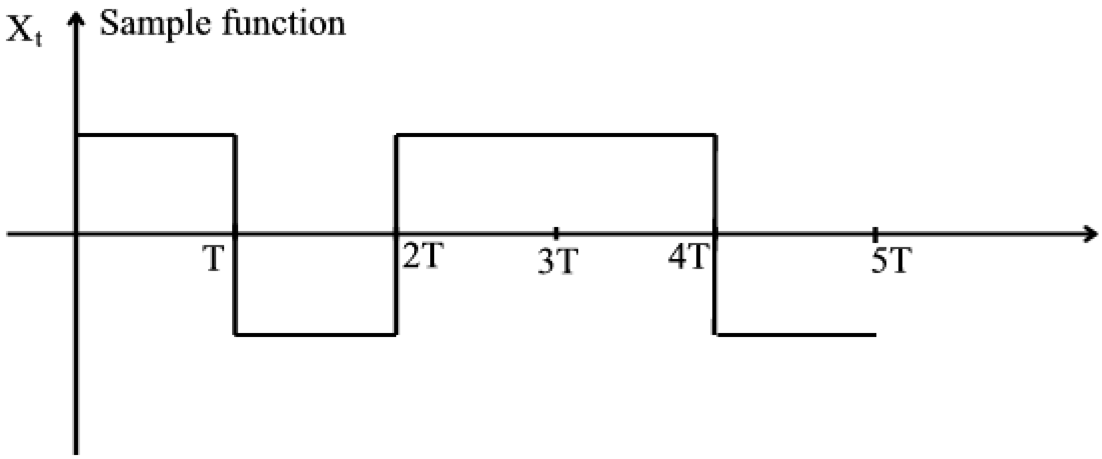
$$= \begin{pmatrix} 0 \\ E_s \end{pmatrix} + \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix}$$

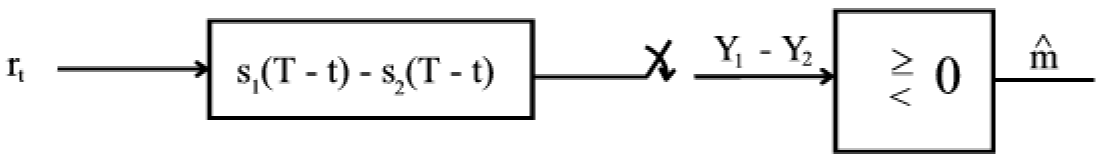
where ν_1 and ν_2 are independent are Gaussian with zero mean and variance $\frac{N_0}{2}E_s$. The analysis is identical to the correlator example.

Equation:

$$P_e = Q\left(\sqrt{\frac{E_s}{N_0}}\right)$$

Note that the maximum likelihood detector decides based on comparing Y_1 and Y_2. If $Y_1 \geq Y_2$ then s_1 was sent; otherwise s_2 was transmitted. For a similar analysis for binary antipodal signals, refer here. See [link] or [link].





Digital Transmission over Baseband Channels

Until this point, we have considered data transmissions over simple additive Gaussian channels that are not time or band limited. In this module we will consider channels that do have bandwidth constraints, and are limited to frequency range around zero (DC). The channel is best modified as $g(t)$ is the impulse response of the baseband channel.

Consider modulated signals $x_t = s_m(t)$ for $0 \leq t \leq T$ for some $m \in \{1, 2, \ldots, M\}$. The channel output is then

Equation:

$$\begin{aligned} r_t &= \int_{-\infty}^{\infty} x_\tau g(t-\tau)\,\mathrm{d}\,\tau + N_t \\ &= \int_{-\infty}^{\infty} S_m(\tau) g(t-\tau)\,\mathrm{d}\,\tau + N_t \end{aligned}$$

The signal contribution in the frequency domain is

Equation:

$$\forall f : \left(\widetilde{S_m}(f) = S_m(f)G(f)\right)$$

The optimum matched filter should match to the filtered signal:

Equation:

$$\forall f : \left(H_m^{\mathrm{opt}}(f) = S_m(f)G(f)e^{(-i)2\pi f t}\right)$$

This filter is indeed **optimum** (i.e., it maximizes signal-to-noise ratio); however, it requires knowledge of the channel impulse response. The signal energy is changed to

Equation:

$$E_{\tilde{s}} = \int_{-\infty}^{\infty} \left(\widetilde{S_m}(f) \right)^2 \mathrm{d}\, f$$

The band limited nature of the channel and the stream of time limited modulated signal create aliasing which is referred to as **intersymbol interference**. We will investigate ISI for a general PAM signaling.

Pulse Amplitude Modulation Through Bandlimited Channel

Consider a PAM system b_{-10},…, b_{-1}, b_0 b_1,…

This implies
Equation:

$$\forall a_n, a_n \in \{\text{M levels of amplitude}\} : \left(x_t = \sum_{n=-\infty}^{\infty} a_n s(t - nT)\right)$$

The received signal is
Equation:

$$\begin{aligned} r_t &= \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} a_n s(t - (\tau - nT)) g(\tau) \,\mathrm{d}\,\tau + N_t \\ &= \sum_{n=-\infty}^{\infty} a_n \int_{-\infty}^{\infty} s(t - (\tau - nT)) g(\tau) \,\mathrm{d}\,\tau + N_t \\ &= \sum_{n=-\infty}^{\infty} a_n \tilde{s}(t - nT) + N_t \end{aligned}$$

Since the signals span a one-dimensional space, one filter matched to $\tilde{s}(t) = \bar{s}g(t)$ is sufficient.

The matched filter's impulse response is
Equation:

$$\forall t : \left(h^{\mathrm{opt}}(t) = \bar{s}g(T - t)\right)$$

The matched filter output is
Equation:

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} a_n \tilde{s}(t - (\tau - nT)) h^{\mathrm{opt}}(\tau) \,\mathrm{d}\,\tau + \nu(t) \\ &= \sum_{n=-\infty}^{\infty} a_n \int_{-\infty}^{\infty} \tilde{s}(t - (\tau - nT)) h^{\mathrm{opt}}(\tau) \,\mathrm{d}\,\tau + \nu(t) \\ &= \sum_{n=-\infty}^{\infty} a_n u(t - nT) + \nu(t) \end{aligned}$$

The decision on the k^{th} symbol is obtained by sampling the MF output at kT:
Equation:

$$y(kT) = \sum_{n=-\infty}^{\infty} a_n u(kT - nT) + \nu(kT)$$

The k^{th} symbol is of interest:
Equation:

$$y(kT) = a_k u(0) + \sum_{n=-\infty}^{\infty} a_n u(kT - nT) + \nu(kT)$$

where $n \neq k$.

Since the channel is bandlimited, it provides memory for the transmission system. The effect of old symbols (possibly even future signals) lingers and affects the performance of the receiver. The effect of ISI can be eliminated or controlled by proper design of **modulation signals** or **precoding** filters at the transmitter, or by **equalizers** or **sequence detectors** at the receiver.

Precoding and Bandlimited Signals

Precoding

The data symbols are manipulated such that
Equation:

$$y_k(kT) = a_k u(0) + \mathrm{ISI} + \nu(kT)$$

Design of Bandlimited Modulation Signals

Recall that modulation signals are
Equation:

$$X_t = \sum_{n=-\infty}^{\infty} a_n s(t - nT)$$

We can design $s(t)$ such that
Equation:

$$u(nT) = \begin{cases} \text{large} & \text{if } \; n = 0 \\ \text{zero or small} & \text{if } \; n \neq 0 \end{cases}$$

where $y(kT) = a_k u(0) + \sum_{n=-\infty}^{\infty} a_n u(kT - nT) + \nu(kT)$ (ISI is the sum term, and once again, $n \neq k$.) Also, $y(nT) = sgh^{\text{opt}}(nT)$ The signal $s(t)$ can be designed to have reduced ISI.

Design Equalizers at the Receiver

Linear equalizers or decision feedback equalizers reduce ISI in the statistic y_t

Maximum Likelihood Sequence Detection

Equation:

$$y(kT) = \sum_{n=-\infty}^{\infty} a_n \left(kT - nT\right) + \nu(k(T))$$

By observing $y(T)$,$y(2T)$,. . . the date symbols are observed frequently. Therefore, ISI can be viewed as diversity to increase performance.

Carrier Phase Modulation

Phase Shift Keying (PSK)

Information is impressed on the phase of the carrier. As data changes from symbol period to symbol period, the phase shifts.
Equation:

$$\forall m, m \in \{1, 2, \ldots, M\} : \left(s_m(t) = AP_T(t)\cos\left(2\pi f_c t + \frac{2\pi(m-1)}{M}\right)\right)$$

Example:
Binary $s_1(t)$ or $s_2(t)$

Representing the Signals

An orthonormal basis to represent the signals is
Equation:

$$\psi_1(t) = \frac{1}{\sqrt{E_s}} AP_T(t)\cos(2\pi f_c t)$$

Equation:

$$\psi_2(t) = \frac{-1}{\sqrt{E_s}} AP_T(t)\sin(2\pi f_c t)$$

The signal
Equation:

$$S_m(t) = AP_T(t)\cos\left(2\pi f_c t + \frac{2\pi(m-1)}{M}\right)$$

Equation:

$$S_m(t) = A\cos\left(\frac{2\pi(m-1)}{M}\right)P_T(t)\cos(2\pi f_c t) - A\sin\left(\frac{2\pi(m-1)}{M}\right)P_T(t)\sin(2\pi f_c t)$$

The signal energy
Equation:

$$\begin{aligned} E_s &= {}_{-\infty}^{\infty} A^2 {P_T}^2(t)\cos^2 \; 2\pi f_c t + \tfrac{2\pi(m-1)}{M} \quad \mathrm{d}\,t \\ &= {}_{0}^{T} A^2 \;\; \tfrac{1}{2} + \tfrac{1}{2}\cos \;\; 4\pi f_c t + \tfrac{4\pi(m-1)}{M} \quad \mathrm{d}\,t \end{aligned}$$

Equation:

$$E_s = \frac{A^2 T}{2} + \frac{1}{2} A^2 \int_0^T \cos\left(4\pi f_c t + \frac{4\pi (m-1)}{M}\right) \mathrm{d}\, t \simeq \frac{A^2 T}{2}$$

(Note that in the above equation, the integral in the last step before the aproximation is very small.) Therefore,
Equation:

$$\psi_1(t) = \sqrt{\frac{2}{T}} P_T(t) \cos(2\pi f_c t)$$

Equation:

$$\psi_2(t) = -\sqrt{\frac{2}{T}} P_T(t) \sin(2\pi f_c t)$$

In general,
Equation:

$$\forall m, m \in \{1, 2, \ldots, M\} : \left(s_m(t) = A P_T(t) \cos\left(2\pi f_c t + \frac{2\pi (m-1)}{M}\right)\right)$$

and $\psi_1(t)$
Equation:

$$\psi_1(t) = \sqrt{\frac{2}{T}} P_T(t) \cos(2\pi f_c t)$$

Equation:

$$\psi_2(t) = \sqrt{\frac{2}{T}} P_T(t) \sin(2\pi f_c t)$$

Equation:

$$s_m = \begin{array}{c} \sqrt{E_s} \cos \frac{2\pi(m-1)}{M} \\ \sqrt{E_s} \sin \frac{2\pi(m-1)}{M} \end{array}$$

Demodulation and Detection

Equation:

$$r_t = s_m(t) + N_t, \text{ for some} m \in \{1, 2, \ldots, M\}$$

We must note that due to phase offset of the oscillator at the transmitter, **phase jitter** or **phase changes** occur because of propagation delay.
Equation:

$$r_t = AP_T(t)\cos\left(2\pi f_c t + \frac{2\pi(m-1)}{M} + \varphi\right) + N_t$$

For binary PSK, the modulation is antipodal, and the optimum receiver in AWGN has average bit-error probability

Equation:

$$\begin{aligned} P_e &= Q\left(\sqrt{\frac{2(E_s)}{N_0}}\right) \\ &= Q\left(A\sqrt{\frac{T}{N_0}}\right) \end{aligned}$$

The receiver where

Equation:

$$r_t = \pm(AP_T(t)\cos(2\pi f_c t + \varphi)) + N_t$$

The statistics

Equation:

$$\begin{aligned} r_1 &= \int_0^T r_t \alpha \cos\left(2\pi f_c t + \varphi\right) \,\mathrm{d}t \\ &= \pm \int_0^T \alpha A \cos(2\pi f_c t + \varphi)\cos\left(2\pi f_c t + \varphi\right) \,\mathrm{d}t + \int_0^T \alpha \cos\left(2\pi f_c t + \varphi\right) N_t \,\mathrm{d}t \end{aligned}$$

Equation:

$$r_1 = \pm\left(\frac{\alpha A}{2}\int_0^T \cos\left(4\pi f_c t + \varphi + \varphi\right) + \cos\left(\varphi - \varphi\right) \,\mathrm{d}t\right) + \eta_1$$

Equation:

$$r_1 = \pm\left(\frac{\alpha A}{2} T\cos\left(\varphi - \varphi\right)\right) + \int_0^T \pm\left(\frac{\alpha A}{2}\cos\left(4\pi f_c t + \varphi + \varphi\right)\right) \mathrm{d}t + \eta_1 \pm \left(\frac{\alpha A T}{2}\cos\left(\varphi - \varphi\right)\right) + \eta_1$$

where $\eta_1 = \alpha \int_0^T N_t \cos\left(\omega_c t + \varphi\right) \,\mathrm{d}t$ is zero mean Gaussian with variance $\simeq \frac{\alpha^2 N_0 T}{4}$.

Therefore,

Equation:

$$\begin{aligned} P_e &= Q\left(\frac{2\frac{\alpha A T}{2}\cos\left(\varphi - \varphi\right)}{2\sqrt{\frac{\alpha^2 N_0 T}{4}}}\right) \\ &= Q\left(\cos\left(\varphi - \varphi\right) A\sqrt{\frac{T}{N_0}}\right) \end{aligned}$$

which is not a function of α and depends strongly on phase accuracy.

Equation:

$$P_e = Q\left(\cos\left(\varphi - \varphi\right)\sqrt{\frac{2E_s}{N_0}}\right)$$

The above result implies that the amplitude of the local oscillator in the correlator structure does not play a role in the performance of the correlation receiver. However, the accuracy of the phase does indeed play a major role. This point can be seen in the following example:

Example:
Equation:

$$x_{t'} = -1^i A \cos\left(-\left(2\pi f_c t'\right) + 2\pi f_c \tau\right)$$

Equation:

$$x_t = -1^i A \cos\left(2\pi f_c t - \left(2\pi f_c \tau' - 2\pi f_c \tau + \theta'\right)\right)$$

Local oscillator should match to phase θ.

Differential Phase Shift Keying

The phase lock loop provides estimates of the phase of the incoming modulated signal. A phase ambiguity of exactly π is a common occurance in many phase lock loop (PLL) implementations.

Therefore it is possible that, $\theta \quad \theta \quad \pi$ without the knowledge of the receiver. Even if there is no noise, if b then b and if b then b .

In the presence of noise, an incorrect decision due to noise may results in a correct final desicion (in binary case, when there is π phase ambiguity with the probability:

Equation:

$$P_e \qquad Q \quad \sqrt{\frac{E_s}{N}}$$

Consider a stream of bits a_n and BPSK modulated signal

Equation:

$$_{n} \quad {}^{a_n} A P_T \;\; t \quad nT \qquad \pi f_c t \quad \theta$$

In differential PSK, the transmitted bits are first encoded $b_n \quad a_n \quad b_n$ with initial symbol (e.g. b) chosen without loss of generality to be either 0 or 1.

Transmitted DPSK signals

Equation:

$$_{n} \quad {}^{b_n} A P_T \;\; t \quad nT \qquad \pi f_c t \quad \theta$$

The decoder can be constructed as

Equation:

$$b_n \quad b_n \quad b_n \quad a_n \quad b_n$$
$$a_n$$
$$a_n$$

If two consecutive bits are detected correctly, if b_n b_n and b_n b_n then

Equation:

$$a_n \quad b_n \quad b_n$$
$$b_n \quad b_n$$
$$a_n \quad b_n \quad b_n$$
$$a_n$$

if b_n b_n and b_n b_n . That is, two consecutive bits are detected incorrectly. Then,

Equation:

$$a_n \quad b_n \quad b_n$$
$$b_n \quad b_n$$
$$b_n \quad b_n$$
$$b_n \quad b_n$$
$$b_n \quad b_n$$
$$a_n$$

If b_n b_n and b_n b_n , that is, one of two consecutive bits is detected in error. In this case there will be an error and the probability of that error for DPSK is

Equation:

$$P_e \quad a_n \quad a_n$$

$$b_n \quad b_n \; b_n \quad b_n \quad b_n \quad b_n \; b_n \quad b_n$$

$$Q \quad \sqrt{\frac{E_s}{N}} \quad Q \quad \sqrt{\frac{E_s}{N}} \quad Q \quad \sqrt{\frac{E_s}{N}}$$

This approximation holds if Q is small.

Carrier Frequency Modulation

Frequency Shift Keying (FSK)

The data is impressed upon the carrier frequency. Therefore, the M different signals are
Equation:

$$s_m(t) = AP_T(t)\cos(2\pi f_c t + 2\pi(m-1)\Delta(f)t + \theta_m)$$

for $m \in \{1, 2, \ldots, M\}$

The M different signals have M different carrier frequencies with possibly different phase angles since the generators of these carrier signals may be different. The carriers are
Equation:

$$f_1 = f_c$$

$$f_2 = f_c + \Delta(f)$$

$$f_M = f_c - M\Delta(f)$$

Thus, the M signals may be designed to be orthogonal to each other.
Equation:

$$\begin{aligned}\langle s_m, s_n\rangle &= \int_0^T A^2\cos(2\pi f_c t + 2\pi(m-1)\Delta(f)t + \theta_m)\cos(2\pi f_c t + 2\pi(n-1)\Delta(f)t + \theta_n)\,\mathrm{d}\,t\\ &= \frac{A^2}{2}\int_0^T \cos(4\pi f_c t + 2\pi(n+m-2)\Delta(f)t + \theta_m + \theta_n)\,\mathrm{d}\,t + \frac{A^2}{2}\int_0^T \cos(2\pi(m-n)\Delta(f)t + \theta_m\\ &= \frac{A^2}{2}\frac{\sin(4\pi f_c T + 2\pi(n+m-2)\Delta(f)T + \theta_m + \theta_n) - \sin(\theta_m + \theta_n)}{4\pi f_c + 2\pi(n+m-2)\Delta(f)} + \frac{A^2}{2}\left(\frac{\sin(2\pi(m-n)\Delta(f)T + \theta_m - \theta_n)}{2\pi(m-n)\Delta(f)} - \frac{\sin(\theta_m - \theta_n)}{2\pi(m-n)\Delta(f)}\right)\end{aligned}$$

If $2f_c T + (n+m-2)\Delta(f)T$ is an integer, and if $(m-n)\Delta(f)T$ is also an integer, then $\langle S_m, S_n\rangle = 0$ if $\Delta(f)T$ is an integer, then $\langle s_m, s_n\rangle \simeq 0$ when f_c is much larger than $\frac{1}{T}$.

In case $\forall m, \theta_m = 0 : (\theta_m = 0)$
Equation:

$$\langle s_m, s_n\rangle \simeq \frac{A^2 T}{2}\operatorname{sinc}(2(m-n)\Delta(f)T)$$

Therefore, the frequency spacing could be as small as $\Delta(f) = \frac{1}{2T}$ since $\operatorname{sinc}(x) = 0$ if $x = \pm(1)$ or $\pm(2)$.

If the signals are designed to be orthogonal then the average probability of error for binary FSK with optimum receiver is
Equation:

$$\bar{P}_e = Q\left(\sqrt{\frac{E_s}{N_0}}\right)$$

in AWGN.

Note that $\operatorname{sinc}(x)$ takes its minimum value not at $x = \pm(1)$ but at $\pm(1.4)$ and the minimum value is -0.216. Therefore if $\Delta(f) = \frac{0.7}{T}$ then

Equation:

$$\bar{P}_e = Q\left(\sqrt{\frac{1.216E_s}{N_0}}\right)$$

which is a gain of $10 \times \log 1.216 \simeq 0.85d\theta$ over orthogonal FSK.

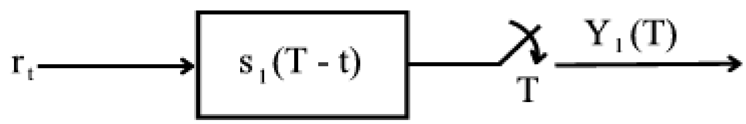
Information Theory and Coding

In the previous chapters, we considered the problem of digital transmission over different channels. Information sources are not often digital, and in fact, many sources are analog. Although many channels are also analog, it is still more efficient to convert analog sources into digital data and transmit over analog channels using digital transmission techniques. There are two reasons why digital transmission could be more efficient and more reliable than analog transmission:

1. Analog sources could be compressed to digital form efficiently.
2. Digital data can be transmitted over noisy channels reliably.

There are several key questions that need to be addressed:

1. How can one model information?
2. How can one quantify information?
3. If information can be measured, does its information quantity relate to how much it can be compressed?
4. Is it possible to determine if a particular channel can handle transmission of a source with a particular information quantity?



Example:
The information content of the following sentences: "Hello, hello, hello." and "There is an exam today." are not the same. Clearly the second one carries more information. The first one can be compressed to "Hello" without much loss of information.

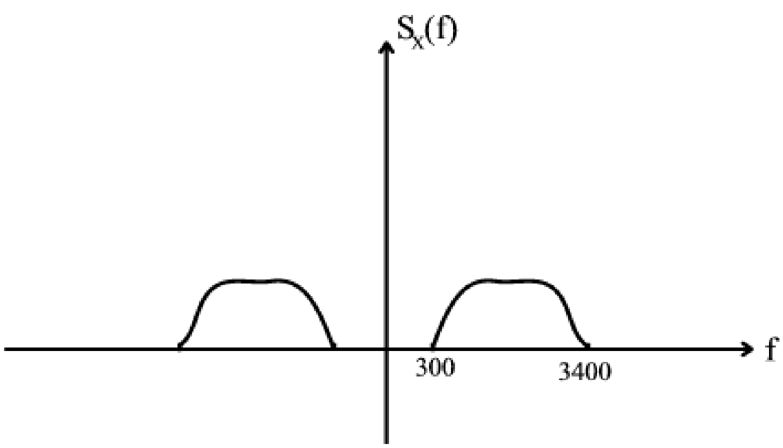
In other modules, we will quantify information and find efficient representation of information (Entropy). We will also quantify how much information can be transmitted through channels, reliably. Channel coding can be used to reduce information rate and increase reliability.

Entropy

Information sources take very different forms. Since the information is not known to the destination, it is then best modeled as a random process, discrete-time or continuous time.

Here are a few examples:

- Digital data source (e.g., a text) can be modeled as a discrete-time and discrete valued random process X_1, X_2, …, where $X_i \in \{A, B, C, D, E, \ldots\}$ with a particular $p_{X_1}(x)$, $p_{X_2}(x)$, …, and a specific $p_{X_1X_2}$, $p_{X_2X_3}$, …, and $p_{X_1X_2X_3}$, $p_{X_2X_3X_4}$, …, etc.
- Video signals can be modeled as a continuous time random process. The power spectral density is bandlimited to around 5 MHz (the value depends on the standards used to raster the frames of image).
- Audio signals can be modeled as a continuous-time random process. It has been demonstrated that the power spectral density of speech signals is bandlimited between 300 Hz and 3400 Hz. For example, the speech signal can be modeled as a Gaussian process with the shown power spectral density over a small observation period.



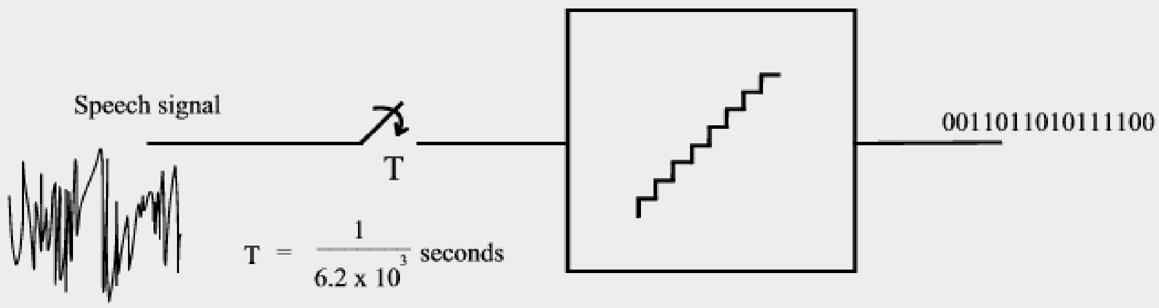
These analog information signals are bandlimited. Therefore, if sampled faster than the Nyquist rate, they can be reconstructed from their sample values.

Example:
A speech signal with bandwidth of 3100 Hz can be sampled at the rate of 6.2 kHz. If the samples are quantized with a 8 level quantizer then the speech signal can be

represented with a binary sequence with the rate of

Equation:

$$
\begin{aligned}
6.2 \times 10^3 \log_2 8 &= 18600 \frac{\text{bits}}{\text{sample}} \frac{\text{samples}}{\text{sec}} \\
&= 18.6 \frac{\text{kbits}}{\text{sec}}
\end{aligned}
$$



The sampled real values can be quantized to create a discrete-time discrete-valued random process. Since any bandlimited analog information signal can be converted to a sequence of discrete random variables, we will continue the discussion only for discrete random variables.

Example:

The random variable $\boldsymbol{x}$ takes the value of 0 with probability 0.9 and the value of 1 with probability 0.1. The statement that $\boldsymbol{x} = 1$ carries more information than the statement that $\boldsymbol{x} = 0$. The reason is that $\boldsymbol{x}$ is expected to be 0, therefore, knowing that $\boldsymbol{x} = 1$ is more surprising news!! An intuitive definition of information measure should be larger when the probability is small.

Example:

The information content in the statement about the temperature and pollution level on July 15th in Chicago should be the sum of the information that July 15th in Chicago was hot and highly polluted since pollution and temperature could be independent.

Equation:

$$I(\text{hot}, \text{high}) = I(\text{hot}) + I(\text{high})$$

An intuitive and meaningful measure of information should have the following properties:

1. Self information should decrease with increasing probability.
2. Self information of two independent events should be their sum.
3. Self information should be a continuous function of the probability.

The only function satisfying the above conditions is the -log of the probability.

Entropy

The entropy (average self information) of a discrete random variable X is a function of its probability mass function and is defined as

Equation:

$$H(X) = -\sum_{i=1}^{N} \mathrm{p}_X(x_i) \log \mathrm{p}_X(x_i)$$

where N is the number of possible values of X and $\mathrm{p}_X(x_i) = \Pr[X = x_i]$. If log is base 2 then the unit of entropy is bits. Entropy is a measure of uncertainty in a random variable and a measure of information it can reveal.

A more basic explanation of entropy is provided in another module.

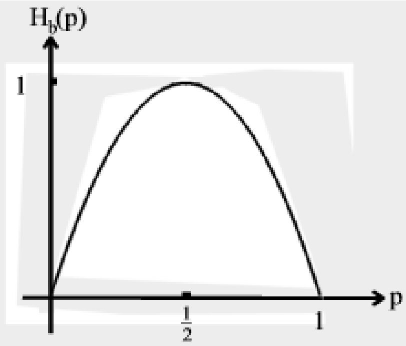
Example:

If a source produces binary information $\{0, 1\}$ with probabilities p and $1 - p$. The entropy of the source is

Equation:

$$H(X) = (-(p \log_2 p)) - (1 - p) \log_2 (1 - p)$$

If $p = 0$ then $H(X) = 0$, if $p = 1$ then $H(X) = 0$, if $p = 1/2$ then $H(X) = 1$ bits. The source has its largest entropy if $p = 1/2$ and the source provides no new information if $p = 0$ or $p = 1$.



Example:
An analog source is modeled as a continuous-time random process with power spectral density bandlimited to the band between 0 and 4000 Hz. The signal is sampled at the Nyquist rate. The sequence of random variables, as a result of sampling, are assumed to be independent. The samples are quantized to 5 levels $\{-2, -1, 0, 1, 2\}$. The probability of the samples taking the quantized values are $\left\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}\right\}$, respectively. The entropy of the random variables are
Equation:

$$\begin{aligned} H(X) &= \left(-\left(\frac{1}{2}\log_2\frac{1}{2}\right)\right) - \frac{1}{4}\log_2\frac{1}{4} - \frac{1}{8}\log_2\frac{1}{8} - \frac{1}{16}\log_2\frac{1}{16} - \frac{1}{16}\log_2\frac{1}{16} \\ &= \frac{1}{2}\log_2 2 + \frac{1}{4}\log_2 4 + \frac{1}{8}\log_2 8 + \frac{1}{16}\log_2 16 + \frac{1}{16}\log_2 16 \\ &= \frac{1}{2} + \frac{1}{2} + \frac{3}{8} + \frac{4}{8} \\ &= \frac{15}{8}\frac{\text{bits}}{\text{sample}} \end{aligned}$$

There are 8000 samples per second. Therefore, the source produces $8000 \times \frac{15}{8} = 15000\frac{\text{bits}}{\text{sec}}$ of information.

Joint Entropy

The joint entropy of two discrete random variables (X, Y) is defined by
Equation:

$$H(X, Y) = -\sum_{ii}\sum_{jj} \mathrm{p}_{X,Y}\ (x_i, y_j)\log \mathrm{p}_{X,Y}\ (x_i, y_j)$$

The joint entropy for a random vector $\boldsymbol{X} = (X_1 X_2 ... X_n)^T$ is defined as
Equation:

$$H(\boldsymbol{X}) = -\sum_{x_1 x_1} \sum_{x_2 x_2} \cdots \sum_{x_n x_n} \mathrm{p}_{\boldsymbol{X}}\ (x_1, x_2, ..., x_n) \log \mathrm{p}_{\boldsymbol{X}}\ (x_1, x_2, ..., x_n)$$

Conditional Entropy

The conditional entropy of the random variable X given the random variable Y is defined by
Equation:

$$H(X|Y) = -\sum_{ii} \sum_{jj} \mathrm{p}_{X,Y}\ (x_i, y_j) \log p_{X|Y}(x_i|y_j)$$

It is easy to show that
Equation:

$$H(\boldsymbol{X}) = H(X_1) + H(X_2|X_1) + \ldots + H(X_n|X_1 X_2 ... X_{\text{n-1}})$$

and
Equation:

$$\begin{aligned} H(X,Y) &= H(Y) + H(X|Y) \\ &= H(X) + H(Y|X) \end{aligned}$$

If X_1, X_2, ..., X_n are mutually independent it is easy to show that
Equation:

$$H(\boldsymbol{X}) = \sum_{i=1}^{n} H(X_i)$$

Entropy Rate

The entropy rate of a stationary discrete-time random process is defined by
Equation:

$$H = \operatorname*{limit}_{n \to \infty}\ H(X_n | X_1 X_2 ... X_n)$$

The limit exists and is equal to
Equation:

$$H = \operatorname{limit}_{n \to \infty} \frac{1}{n} H(X_1, X_2, ..., X_n)$$

The entropy rate is a measure of the uncertainty of information content per output symbol of the source.

Entropy is closely tied to source coding. The extent to which a source can be compressed is related to its entropy. In 1948, Claude E. Shannon introduced a theorem which related the entropy to the number of bits per second required to represent a source without much loss.

Source Coding

As mentioned earlier, how much a source can be compressed should be related to its entropy. In 1948, Claude E. Shannon introduced three theorems and developed very rigorous mathematics for digital communications. In one of the three theorems, Shannon relates entropy to the minimum number of bits per second required to represent a source without much loss (or distortion).

Consider a source that is modeled by a discrete-time and discrete-valued random process X_1, X_2, ..., X_n, ... where $x_i \in \{a_1, a_2, \ldots, a_N\}$ and define $p_{X_i}(x_i = a_j) = p_j$ for $j = 1\ 2\ \ldots\ N$, where it is assumed that X_1, X_2,... X_n are mutually independent and identically distributed.

Consider a sequence of length n

Equation:

$$= \begin{matrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{matrix}$$

The symbol a_1 can occur with probability p_1. Therefore, in a sequence of length n, on the average, a_1 will appear np_1 times with high probabilities if n is very large.

Therefore,

Equation:

$$P(\quad = \quad) = p_{X_1}(x_1) p_{X_2}(x_2) \ldots p_{X_n}(x_n)$$

Equation:

$$P(\quad = \quad) \simeq {p_1}^{np_1} {p_2}^{np_2} \ldots {p_N}^{np_N} = \overset{N}{\underset{i \quad 1}{}} {p_i}^{np_i}$$

where $p_i = P(X_j = a_i)$ for all j and for all i.

A typical sequence may look like
Equation:

$$
= \begin{matrix} a_2 \\ \vdots \\ a_1 \\ a_N \\ a_2 \\ a_5 \\ \vdots \\ a_1 \\ \vdots \\ a_N \\ a_6 \end{matrix}
$$

where a_i appears np_i times with large probability. This is referred to as a **typical sequence**. The probability of being a typical sequence is
Equation:

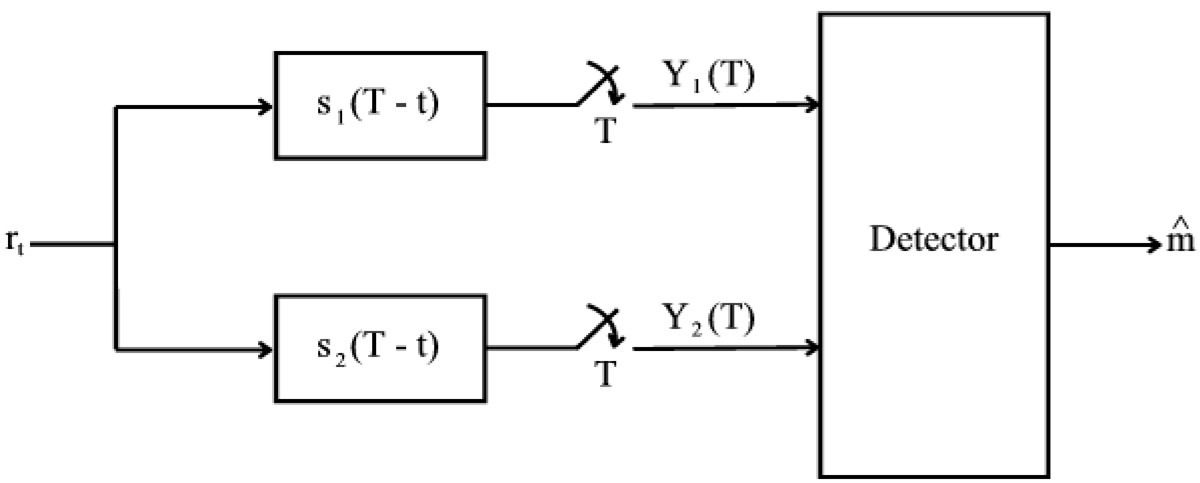
$$
\begin{aligned}
P(\quad = \quad) \simeq \quad {}_{i\ \ 1}^{N} \, {p_i}^{np_i} &= \quad {}_{i\ \ 1}^{N} \; 2^{\log_2 p_i \quad np_i} \\
&= \quad {}_{i\ \ 1}^{N} \, 2^{np_i \log_2 p_i} \\
&= 2^{n \quad {}_{i\ \ 1}^{N} p_i \log_2 p_i} \\
&= 2^{-(nH(X))}
\end{aligned}
$$

where $H(X)$ is the entropy of the random variables X_1, X_2,…, X_n.

For large n, almost all the output sequences of length n of the source are equally probably with probability $\simeq 2^{-(nH(X))}$. These are typical sequences. The probability of nontypical sequences are negligible. There are N^n different sequences of length n with alphabet of size N. The probability of typical sequences is almost 1.

Equation:

$$\overset{\#\text{ of typical seq.}}{\underset{k\quad 1}{}}\; 2^{-(nH(X))} = 1$$



Example:

Consider a source with alphabet {A,B,C,D} with probabilities { $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{8}$ }. Assume X_1, X_2,..., X_8 is an independent and identically distributed sequence with $X_i \in \{A, B, C, D\}$ with the above probabilities.

Equation:

$$
\begin{aligned}
H(X) &= -\tfrac{1}{2}\log_2\tfrac{1}{2} - \tfrac{1}{4}\log_2\tfrac{1}{4} - \tfrac{1}{8}\log_2\tfrac{1}{8} - \tfrac{1}{8}\log_2\tfrac{1}{8} \\
&= \tfrac{1}{2} + \tfrac{2}{4} + \tfrac{3}{8} + \tfrac{3}{8} \\
&= \tfrac{4+4+6}{8} \\
&= \tfrac{14}{8}
\end{aligned}
$$

The number of typical sequences of length 8

Equation:

$$2^{8\times\frac{14}{8}} = 2^{14}$$

The number of nontypical sequences

$4^8 - 2^{14} = 2^{16} - 2^{14} = 2^{14}\left(4-1\right) = 3\times 2^{14}$

Examples of typical sequences include those with A appearing $8\times\frac{1}{2}=4$ times, B appearing $8\times\frac{1}{4}=2$ times, etc. {A,D,B,B,A,A,C,A}, {A,A,A,A,C,D,B,B} and much more.

Examples of nontypical sequences of length 8: {D,D,B,C,C,A,B,D}, {C,C,C,C,C,B,C,C} and much more. Indeed, these definitions and arguments are valid when n is very large. The probability of a source output to be in the set of typical sequences is 1 when $n\rightarrow\infty$. The probability of a source output to be in the set of nontypical sequences approaches 0 as $n\rightarrow\infty$.

The essence of source coding or data compression is that as $n\rightarrow\infty$, nontypical sequences never appear as the output of the source. Therefore, one only needs to be able to represent typical sequences as binary codes and ignore nontypical sequences. Since there are only $2^{nH(X)}$ typical sequences of length n, it takes $nH(X)$ bits to represent them on the average. On the average it takes $H(X)$ bits per source output to represent a simple source that produces independent and identically distributed outputs.

Theorem

Shannon's Source-Coding

A source that produced independent and identically distributed random variables with entropy H can be encoded with arbitrarily small error

probability at any rate R in bits per source output if $R \geq H$. Conversely, if $R < H$, the error probability will be bounded away from zero, independent of the complexity of coder and decoder.

The source coding theorem proves existence of source coding techniques that achieve rates close to the entropy but does not provide any algorithms or ways to construct such codes.

If the source is not i.i.d. (independent and identically distributed), but it is stationary with memory, then a similar theorem applies with the entropy $H(X)$ replaced with the entropy rate $H = \lim_{n\to\infty} H(X_n|X_1X_2\ldots X_{\text{n-1}})$

In the case of a source with memory, the more the source produces outputs the more one knows about the source and the more one can compress.

Example:
The English language has 26 letters, with space it becomes an alphabet of size 27. If modeled as a memoryless source (no dependency between letters in a word) then the entropy is $H(X) = 4.03$ bits/letter.
If the dependency between letters in a text is captured in a model the entropy rate can be derived to be $H = 1.3$ bits/letter. Note that a non-information theoretic representation of a text may require 5 bits/letter since 2^5 is the closest power of 2 to 27. Shannon's results indicate that there may be a compression algorithm with the rate of 1.3 bits/letter.

Although Shannon's results are not constructive, there are a number of source coding algorithms for discrete time discrete valued sources that come close to Shannon's bound. One such algorithm is the Huffman source coding algorithm. Another is the Lempel and Ziv algorithm.

Huffman codes and Lempel and Ziv apply to compression problems where the source produces discrete time and discrete valued outputs. For cases where the source is analog there are powerful compression algorithms that specify all the steps from sampling, quantizations, and binary

representation. These are referred to as waveform coders. JPEG, MPEG, vocoders are a few examples for image, video, and voice, respectively.

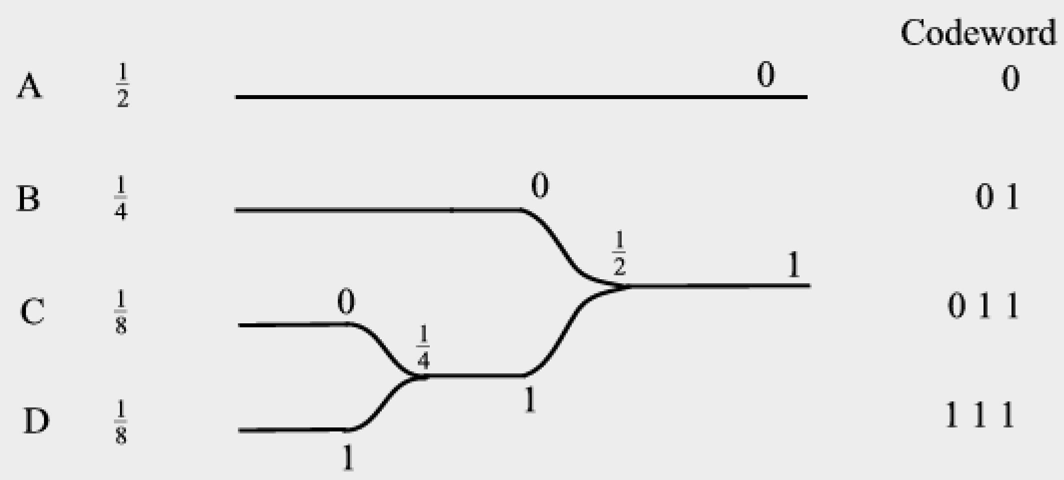
Huffman Coding

One particular source coding algorithm is the Huffman encoding algorithm. It is a source coding algorithm which approaches, and sometimes achieves, Shannon's bound for source compression. A brief discussion of the algorithm is also given in another module.

Huffman encoding algorithm

1. Sort source outputs in decreasing order of their probabilities
2. Merge the two least-probable outputs into a single output whose probability is the sum of the corresponding probabilities.
3. If the number of remaining outputs is more than 2, then go to step 1.
4. Arbitrarily assign 0 and 1 as codewords for the two remaining outputs.
5. If an output is the result of the merger of two outputs in a preceding step, append the current codeword with a 0 and a 1 to obtain the codeword the the preceding outputs and repeat step 5. If no output is preceded by another output in a preceding step, then stop.

Example:
$X \in \{A, B, C, D\}$ with probabilities $\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\}$



Average length $= \frac{1}{2}1 + \frac{1}{4}2 + \frac{1}{8}3 + \frac{1}{8}3 = \frac{14}{8}$. As you may recall, the entropy of the source was also $H(X) = \frac{14}{8}$. In this case, the Huffman code achieves the lower bound of $\frac{14}{8} \frac{\text{bits}}{\text{output}}$.

In general, we can define average code length as

Equation:

$$\overline{\ell} = \sum_{x \in X} \mathrm{p}_X\ (x)\ell(x)$$

where X is the set of possible values of x.

It is not very hard to show that

Equation:

$$H(X) \geq \overline{\ell} > H(X) + 1$$

For compressing single source output at a time, Huffman codes provide nearly optimum code lengths.

The drawbacks of Huffman coding

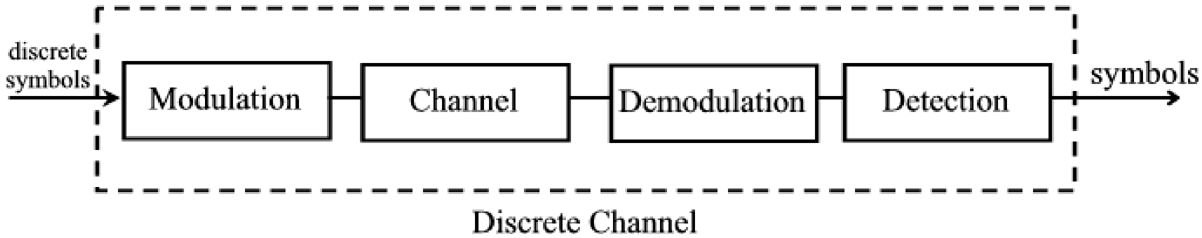
1. Codes are variable length.
2. The algorithm requires the knowledge of the probabilities, $\mathrm{p}_X\ (x)$ for all $x \in X$.

Another powerful source coder that does not have the above shortcomings is Lempel and Ziv.

Channel Capacity

In the previous section, we discussed information sources and quantified information. We also discussed how to represent (and compress) information sources in binary symbols in an efficient manner. In this section, we consider channels and will find out how much information can be sent through the channel reliably.

We will first consider simple channels where the input is a discrete random variable and the output is also a discrete random variable. These discrete channels could represent analog channels with modulation and demodulation and detection.



Let us denote the input sequence to the channel as
Equation:

$$\boldsymbol{X}$$

where a discrete symbol set or input alphabet.

The channel output
Equation:

$$Y$$

where a discrete symbol set or output alphabet.

The statistical properties of a channel are determined if one finds $Y X \, y \, x$ for all y and for all x. A discrete channel is called a **discrete memoryless channel** if

Equation:

$$Y X \, y \, x$$

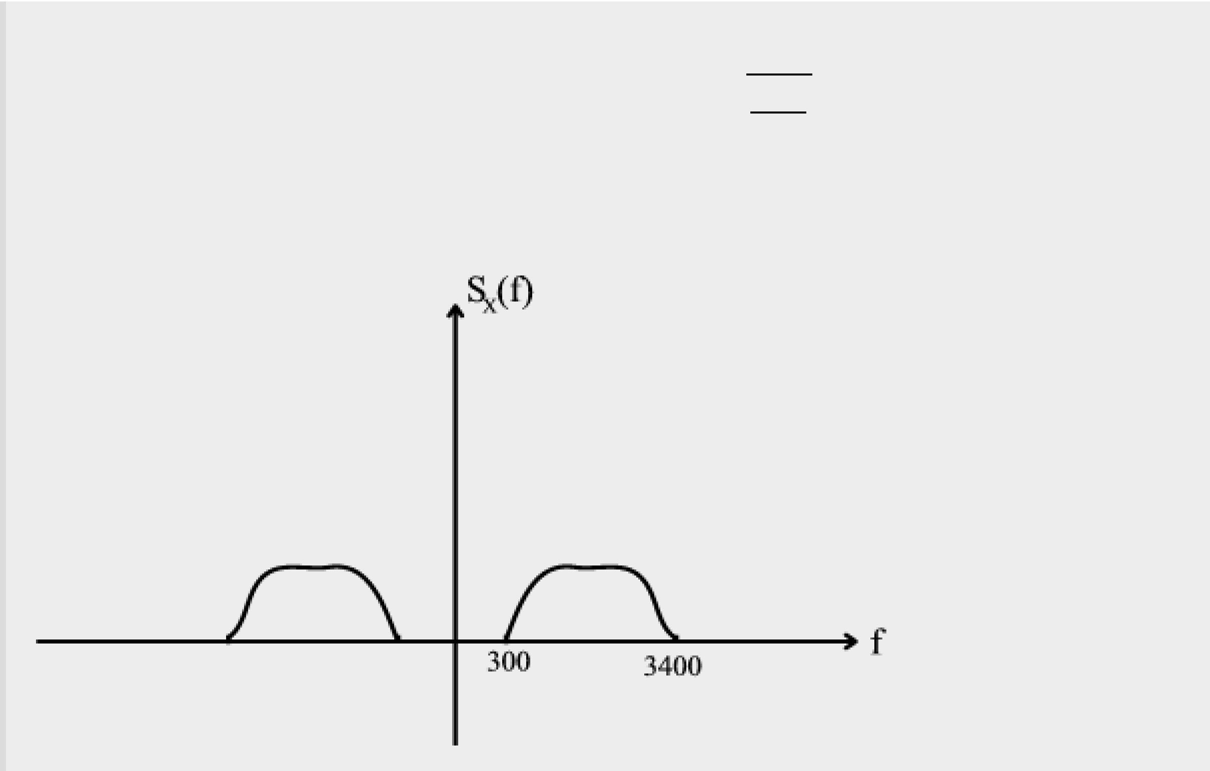
for all y and for all x .

Example:

A binary symmetric channel (BSC) is a discrete memoryless channel with binary input and binary output and .

As an example, a white Gaussian channel with antipodal signaling and matched filter receiver has probability of error of . Since the error is symmetric with respect to the transmitted bit, then

Equation:



It is interesting to note that every time a BSC is used one bit is sent across the channel with probability of error of . The question is how much information or how many bits can be sent per channel use, reliably. Before we consider the above question a few definitions are essential. These are discussed in mutual information.

Mutual Information

Recall that

Equation:

$$H(X,Y) = -\sum_{xx}\sum_{yy} \mathrm{p}_{X,Y}\ (x,y)\log \mathrm{p}_{X,Y}\ (x,y)$$

Equation:

$$H(Y) + H(X|Y) = H(X) + H(Y|X)$$

Mutual Information

The mutual information between two discrete random variables is denoted by $\mathscr{I}(X;Y)$ and defined as

Equation:

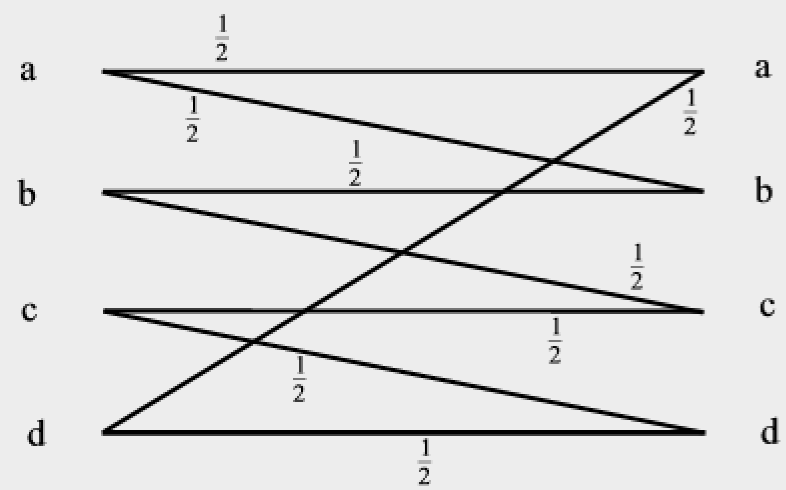
$$\mathscr{I}(X;Y) = H(X) - H(X|Y)$$

Mutual information is a useful concept to measure the amount of information shared between input and output of noisy channels.

In our previous discussions it became clear that when the channel is noisy there may not be reliable communications. Therefore, the limiting factor could very well be reliability when one considers noisy channels. Claude E. Shannon in 1948 changed this paradigm and stated a theorem that presents the rate (speed of communication) as the limiting factor as opposed to reliability.

Example:

Consider a discrete memoryless channel with four possible inputs and outputs.

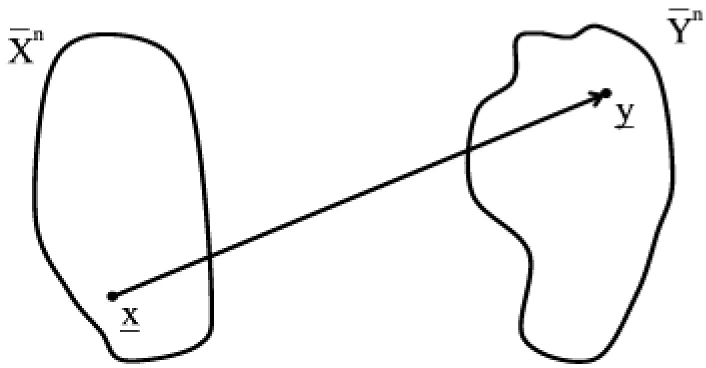


Every time the channel is used, one of the four symbols will be transmitted. Therefore, 2 bits are sent per channel use. The system, however, is very unreliable. For example, if "a" is received, the receiver can not determine, reliably, if "a" was transmitted or "d". However, if the transmitter and receiver agree to only use symbols "a" and "c" and never use "b" and "d", then the transmission will always be reliable, but 1 bit is sent per channel use. Therefore, the rate of transmission was the limiting factor and not reliability.

This is the essence of Shannon's noisy channel coding theorem, i.e., using only those inputs whose corresponding outputs are disjoint (e.g., far apart). The concept is appealing, but does not seem possible with binary channels since the input is either zero or one. It may work if one considers a vector of binary inputs referred to as the extension channel.

$$\mathbf{X\ input\ vector} = \begin{matrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{matrix} \in X^n = \{0,1\}^n$$

$$\textbf{Y output vector} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \in Y^n = \{0,1\}^n$$



This module provides a description of the basic information necessary to understand Shannon's Noisy Channel Coding Theorem. However, for additional information on typical sequences, please refer to Typical Sequences.

Typical Sequences

If the binary symmetric channel has crossover probability ε then if $\boldsymbol{x}$ is transmitted then by the Law of Large Numbers the output $\boldsymbol{y}$ is different from $\boldsymbol{x}$ in $n\varepsilon$ places if n is very large.
Equation:

$$d_H(\boldsymbol{x}, \boldsymbol{y}) \simeq n\varepsilon$$

The number of sequences of length n that are different from $\boldsymbol{x}$ of length n at $n\varepsilon$ is
Equation:

$$\binom{n}{n\varepsilon} = \frac{n!}{(n\varepsilon)!\,(n - n\varepsilon)!}$$

Example:
$\boldsymbol{x} = (000)^T$ and $\varepsilon = \frac{1}{3}$ and $n\varepsilon = 3 \times \frac{1}{3}$. The number of output sequences different from $\boldsymbol{x}$ by one element: $\frac{3!}{1!2!} = \frac{3\times2\times1}{1\times2} = 3$ given by $(101)^T$, $(011)^T$, and $(000)^T$.

Using Stirling's approximation
Equation:

$$n! \simeq n^n e^{-n}\sqrt{2\pi n}$$

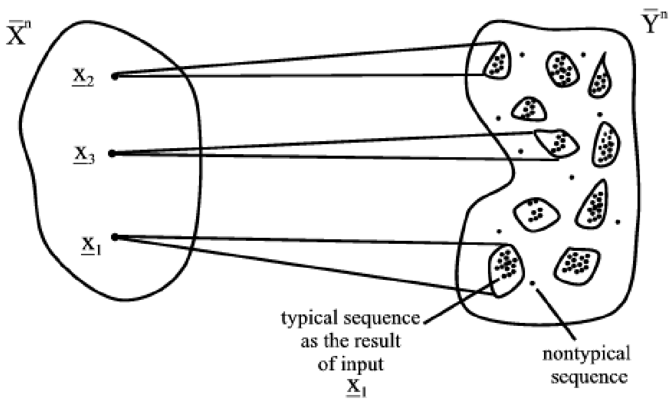
we can approximate
Equation:

$$\binom{n}{n\varepsilon} \simeq 2^{\mathrm{n}((-(\varepsilon\log_2\varepsilon)) - (1-\varepsilon)\log_2(1-\varepsilon))} = 2^{nH_b(\varepsilon)}$$

where $H_b(\varepsilon) \equiv (-(\varepsilon\log_2\varepsilon)) - (1-\varepsilon)\log_2(1-\varepsilon)$ is the entropy of a binary memoryless source. For any $\boldsymbol{x}$ there are $2^{nH_b(\varepsilon)}$ highly probable outputs that correspond to this input.

Consider the output vector $\boldsymbol{Y}$ as a very long random vector with entropy $nH(Y)$. As discussed earlier, the number of typical sequences (or highly probably) is roughly $2^{nH(Y)}$. Therefore, 2^n is the total number of binary sequences, $2^{nH(Y)}$ is the number of typical sequences, and $2^{nH_b(\varepsilon)}$ is the number of elements in a group of possible outputs for one input vector. The maximum number of input sequences that produce nonoverlapping output sequences
Equation:

$$\begin{aligned} M &= \frac{2^{nH(Y)}}{2^{nH_b(\varepsilon)}} \\ &= 2^{n(H(Y) - H_b(\varepsilon))} \end{aligned}$$



The number of distinguishable input sequences of length n is
Equation:

$$2^{n(H(Y)-H_b(\varepsilon))}$$

The number of information bits that can be sent across the channel reliably per n channel uses $n\,(H(Y)-H_b(\varepsilon))$ The maximum reliable transmission rate per channel use
Equation:

$$\begin{aligned} R &= \frac{\log_2 M}{n} \\ &= \frac{n(H(Y)-H_b(\varepsilon))}{n} \\ &= H(Y)-H_b(\varepsilon) \end{aligned}$$

The maximum rate can be increased by increasing $H(Y)$. Note that $H_b(\varepsilon)$ is only a function of the crossover probability and can not be minimized any further.

The entropy of the channel output is the entropy of a binary random variable. If the input is chosen to be uniformly distributed with $p_X(0)=p_X(1)=\frac{1}{2}$.

Then
Equation:

$$\begin{aligned} p_Y(0) &= 1p_X(0)+\varepsilon p_X(1) \\ &= \frac{1}{2} \end{aligned}$$

and
Equation:

$$\begin{aligned} p_Y(1) &= 1p_X(1)+\varepsilon p_X(0) \\ &= \frac{1}{2} \end{aligned}$$

Then, $H(Y)$ takes its maximum value of 1. Resulting in a maximum rate $R = \quad - H_b(\varepsilon)$ when $p_X(0)=p_X(1)=\frac{1}{2}$. This result says that ordinarily one bit is transmitted across a BSC with reliability

$1-\varepsilon$. If one needs to have probability of error to reach zero then one should reduce transmission of information to $1-H_b(\varepsilon)$ and add redundancy.

Recall that for Binary Symmetric Channels (BSC)

Equation:

$$\begin{aligned} H(Y|X) &= p_x(0)H(Y|X=0)+p_x(1)H(Y|X=1) \\ &= p_x(0)\left(-\left((1-\varepsilon)\log_2(1-\varepsilon)-\varepsilon\log_2\varepsilon\right)\right)+p_x(1)\left(-\left((1-\varepsilon)\log_2(1-\varepsilon)-\varepsilon\log_2\varepsilon\right)\right) \\ &= \left(-\left((1-\varepsilon)\log_2(1-\varepsilon)\right)\right)-\varepsilon\log_2\varepsilon \\ &= H_b(\varepsilon) \end{aligned}$$

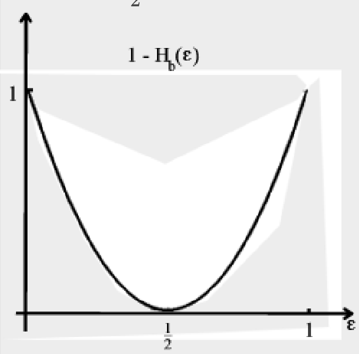
Therefore, the maximum rate indeed was

Equation:

$$\begin{aligned} R &= H(Y)-H(Y|X) \\ &= \quad (X;Y) \end{aligned}$$

Example:

The maximum reliable rate for a BSC is $1-H_b(\varepsilon)$. The rate is 1 when $\varepsilon=0$ or $\varepsilon=1$. The rate is 0 when $\varepsilon=\frac{1}{2}$



This module provides background information necessary for an understanding of [Shannon's Noisy Channel Coding Theorem](). It is also closely related to material presented in [Mutual Information]().

Shannon's Noisy Channel Coding Theorem

It is highly recommended that the information presented in Mutual Information and in Typical Sequences be reviewed before proceeding with this document. An introductory module on the theorem is available at Noisy Channel Theorems .

Theorem
Shannon's Noisy Channel Coding

The capacity of a discrete-memoryless channel is given by
Equation:

$$C = \max_{p_X(x)} \{\mathscr{I}(X;Y) | \, p_X(x)\}$$

where $\mathscr{I}(X;Y)$ is the mutual information between the channel input X and the output Y. If the transmission rate R is less than C, then for any $\varepsilon > 0$ there exists a code with block length n large enough whose error probability is less than ε. If $R > C$, the error probability of any code with any block length is bounded away from zero.

Example:
If we have a binary symmetric channel with cross over probability 0.1, then the capacity $C \simeq 0.5$ bits per transmission. Therefore, it is possible to send 0.4 bits per channel through the channel reliably. This means that we can take 400 information bits and map them into a code of length 1000 bits. Then the whole code can be transmitted over the channels. One hundred of those bits may be detected incorrectly but the 400 information bits may be decoded correctly.

Before we consider continuous-time additive white Gaussian channels, let's concentrate on discrete-time Gaussian channels
Equation:

$$Y_i = X_i + \eta_i$$

where the X_i's are information bearing random variables and η_i is a Gaussian random variable with variance σ_η^2. The input X_i's are constrained to have power less than P

Equation:

$$\frac{1}{n} \sum_{i \quad 1}^{n} X_i{}^2 \leq P$$

Consider an output block of size n

Equation:

$$= \quad +$$

For large n, by the Law of Large Numbers,

Equation:

$$\frac{1}{n} \sum_{i \quad 1}^{n} \eta_i{}^2 = \frac{1}{n} \sum_{i \quad 1}^{n} (|y_i - x_i|)^2 \leq \sigma_\eta{}^2$$

This indicates that with large probability as n approaches infinity, will be located in an n-dimensional sphere of radius $\sqrt{n\sigma_\eta{}^2}$ centered about since $(| \quad - \quad |)^2 \leq n\sigma_\eta{}^2$

On the other hand since X_i's are power constrained and η_i and X_i's are independent

Equation:

$$\frac{1}{n} \sum_{i \quad 1}^{n} y_i{}^2 \leq P + \sigma_\eta{}^2$$

Equation:

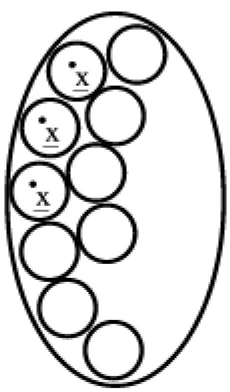
$$| \quad | \leq n \quad P + {\sigma_\eta}^2$$

This mean is in a sphere of radius $\overline{n\left(P + {\sigma_\eta}^2\right)}$ centered around the origin.

How many 's can we transmit to have nonoverlapping spheres in the output domain? The question is how many spheres of radius $\overline{n{\sigma_\eta}^2}$ fit in a sphere of radius $\overline{n\left(P + {\sigma_\eta}^2\right)}$.

Equation:

$$\begin{aligned} M &= \frac{\overline{n({\sigma_\eta}^2+P)}^{\;n}}{\overline{n{\sigma_\eta}^2}^{\;n}} \\ &= 1 + \frac{P}{{\sigma_\eta}^2}^{\;\frac{n}{2}} \end{aligned}$$



Exercise:

Problem:

How many bits of information can one send in n uses of the channel?

Solution:
Equation:

$$\log_2 \quad 1+\frac{P}{{\sigma_\eta}^2} \quad {}^{\frac{n}{2}}$$

The capacity of a discrete-time Gaussian channel $C=\frac{1}{2}\log_2 \quad 1+\frac{P}{{\sigma_\eta}^2}$ bits per channel use.

When the channel is a continuous-time, bandlimited, additive white Gaussian with noise power spectral density $\frac{N_0}{2}$ and input power constraint P and bandwidth W. The system can be sampled at the Nyquist rate to provide power per sample P and noise power

Equation:

$$\begin{aligned} {\sigma_\eta}^2 &= {}_{-W}^{W} \frac{N_0}{2}\, d f \\ &= W N_0 \end{aligned}$$

The channel capacity $\frac{1}{2}\log_2 \quad 1+\frac{P}{N_0 W}$ bits per transmission. Since the sampling rate is $2W$, then

Equation:

$$C=\frac{2W}{2}\log_2 \quad 1+\frac{P}{N_0 W} \quad \text{bits/trans. x trans./sec}$$

Equation:

$$C=W\log_2 \quad 1+\frac{P}{N_0 W} \quad \frac{\text{bits}}{\text{sec}}$$

Example:

The capacity of the voice band of a telephone channel can be determined using the Gaussian model. The bandwidth is 3000 Hz and the signal to

noise ratio is often 30 dB. Therefore,

Equation:

$$C = 3000 \log_2 (1 + 1000) \simeq 30000 \frac{\text{bits}}{\text{sec}}$$

One should not expect to design modems faster than 30 Kbs using this model of telephone channels. It is also interesting to note that since the signal to noise ratio is large, we are expecting to transmit 10 bits/second/Hertz across telephone channels.

Channel Coding

Channel coding is a viable method to reduce information rate through the channel and increase reliability. This goal is achieved by adding redundancy to the information symbol vector resulting in a longer coded vector of symbols that are distinguishable at the output of the channel. Another brief explanation of channel coding is offered in Channel Coding and the Repetition Code. We consider only two classes of codes, block codes and convolutional codes.

Block codes

The information sequence is divided into blocks of length k. Each block is mapped into channel inputs of length n. The mapping is independent from previous blocks, that is, there is no memory from one block to another.

Example:
$k=2$ and $n=5$
Equation:

$$00 \to 00000$$

Equation:

$$01 \to 10100$$

Equation:

$$10 \to 01111$$

Equation:

$$11 \to 11011$$

information sequence ⇒ codeword (channel input)

A binary block code is completely defined by 2^k binary sequences of length n called codewords.
Equation:

$$= \{c_1, c_2, \ldots, c_{2^k}\}$$

Equation:

$$c_i \in \{0, 1\}^n$$

There are three key questions,

1. How can one find "good" codewords?
2. How can one systematically map information sequences into codewords?
3. How can one systematically find the corresponding information sequences from a codeword, i.e., how can we decode?

These can be done if we concentrate on linear codes and utilize finite field algebra.

A block code is linear if $_i \in$ and $_j \in$ implies $_i \oplus {}_j \in$ where $\oplus$ is an elementwise modulo 2 addition.

Hamming distance is a useful measure of codeword properties
Equation:

$$d_H({}_i, {}_j) = \# \text{ of places that they are different}$$

Denote the codeword for information sequence $e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$ by g_1 and $e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$ by g_2,…, and $e_k = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$ by g_k. Then any information sequence can be expressed as

Equation:

$$= \begin{pmatrix} u_1 \\ \vdots \\ u_k \end{pmatrix} = \sum_{i=1}^{k} u_i e_i$$

and the corresponding codeword could be

Equation:

$$= \sum_{i=1}^{k} u_i g_i$$

Therefore

Equation:

$$ = G$$

with $= \{0,1\}^n$ and $\in \{0,1\}^k$ where $G = \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_k \end{pmatrix}$, a kxn matrix and all operations are modulo 2.

Example:
In [link] with
Equation:

$$00 \rightarrow 00000$$

Equation:

$$01 \rightarrow 10100$$

Equation:

$$10 \rightarrow 01111$$

Equation:

$$11 \rightarrow 11011$$

$g_1 = (01111)^T$ and $g_2 = (10100)^T$ and $G = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \end{pmatrix}$

Additional information about coding efficiency and error are provided in Block Channel Coding.

Examples of good linear codes include Hamming codes, BCH codes, Reed-Solomon codes, and many more. The rate of these codes is defined as $\frac{k}{n}$ and these codes have different error correction and error detection properties.

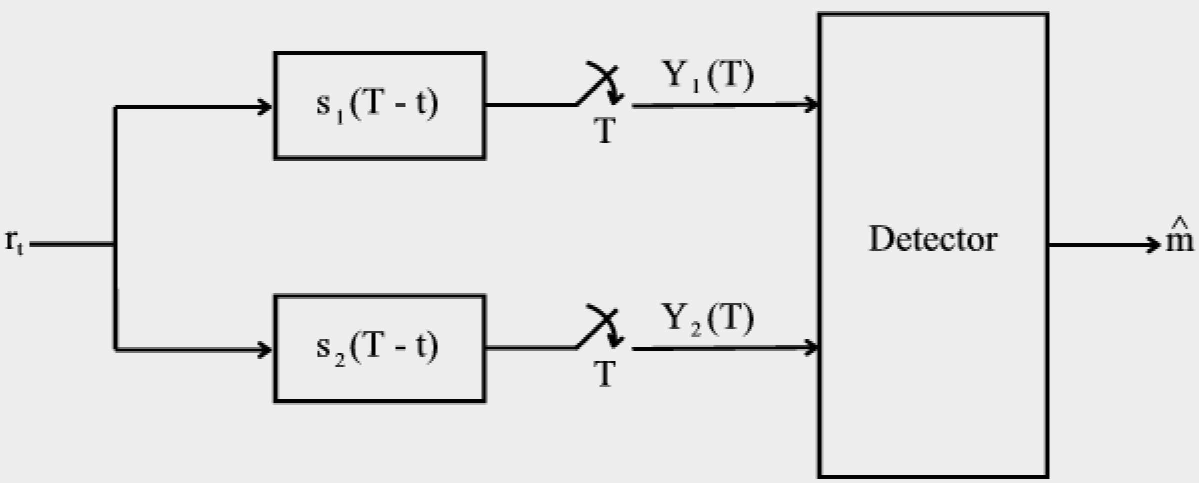
Convolutional Codes

Convolutional codes are one type of code used for channel coding. Another type of code used is block coding.

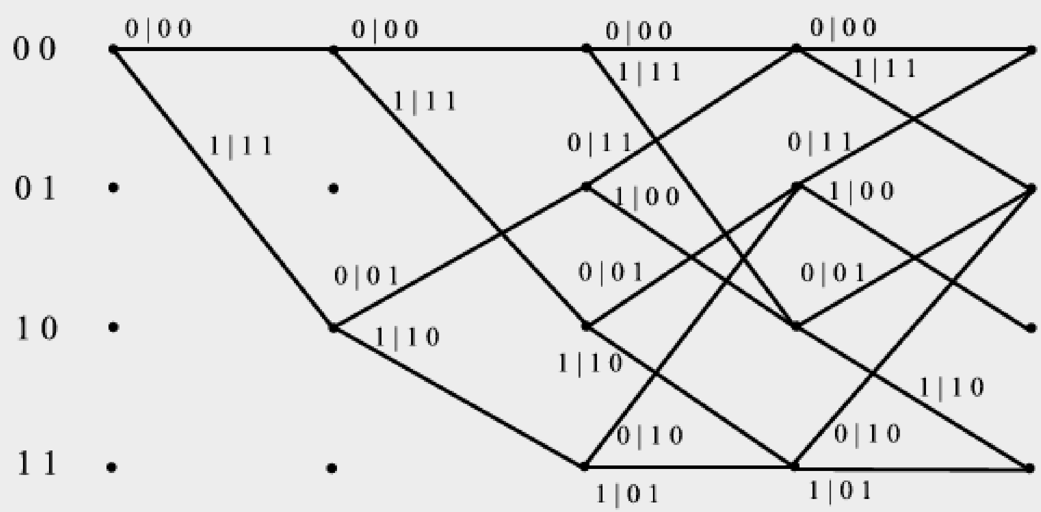
Convolutional codes

In convolutional codes, each block of bits is mapped into a block of bits but these bits are not only determined by the present information bits but also by the previous information bits. This dependence can be captured by a finite state machine.

Example:
A rate – convolutional coder , with memory length 2 and constraint length 3.



Since the length of the shift register is 2, there are 4 different rates. The behavior of the convolutional coder can be captured by a 4 state machine.
States: `00, 01, 10, 11`,
For example, arrival of information bit `0` transitions from state `10` to state `01`.
The encoding and the decoding process can be realized in trellis structure.



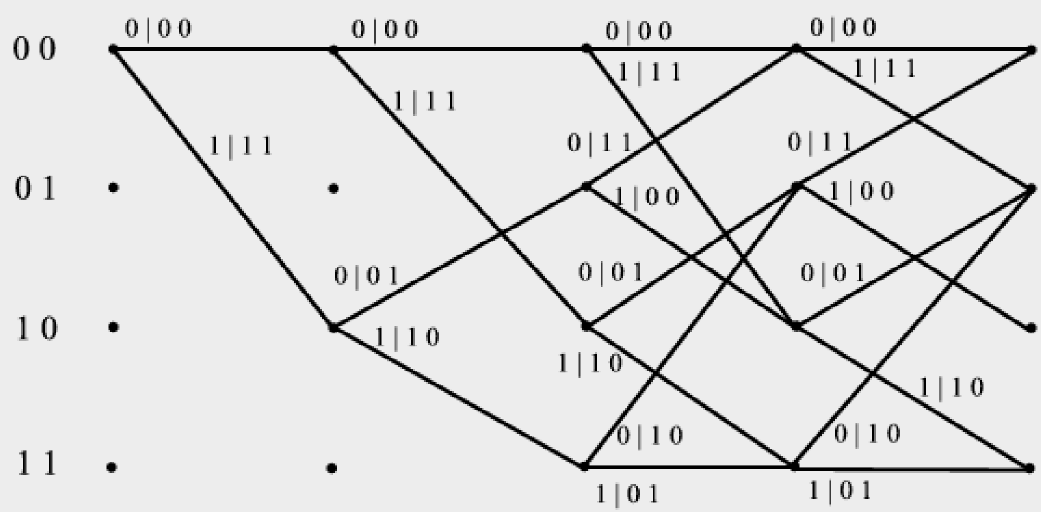
If the input sequence is

```
1 1 0 0
```

the output sequence would be

```
11 10 10 11
```

The transmitted codeword is then `11 10 10 11`. If there is one error on the channel `11 00 10 11`



Starting from state `00` the Hamming distance between the possible paths and the received sequence is measured. At the end, the path with minimum distance to the received sequence is chosen as the correct trellis path. The information sequence will then be determined.

Convolutional coding lends itself to very efficient trellis based encoding and decoding. They are very practical and powerful codes.

Homework 1 of Elec 430

Elec 430 homework set 1. Rice University Department of Electrical and Computer Engineering.

Exercise:

Problem:

The current I in a semiconductor diode is related to the voltage V by the relation $I = e^{V} - 1$. If V is a random variable with density function $f_V(x) = \frac{1}{2} e^{-|x|}$ for $-\infty < x < \infty$, find $f_I(y)$; the density function of I.

Exercise:

Problem:

Show that if $AB = \{\}$ then $\Pr[A] \leq \Pr[B^c]$

Show that for any A, B, C we have
$\Pr[A \cup B \cup C] = \Pr[A] + \Pr[B] + \Pr[C] - \Pr[A \cap B] - \Pr[A \cap C] - \Pr[B \cap C] + \Pr[A \cap B \cap C]$

Show that if A and B are independent the $\Pr[A \cap B^c] = \Pr[A] \Pr[B^c]$ which means A and B^c are also independent.

Exercise:

Problem:

Suppose X is a discrete random variable taking values $\{0, 1, 2, \ldots, n\}$ with the following probability mass function $p_X(k) = \begin{cases} \frac{n!}{k!(n-k)!} \theta^k (1-\theta)^{n-k} & \text{if } k = \{0, 1, 2, \ldots, n\} \\ 0 & \text{otherwise} \end{cases}$ with parameter $\theta \in [0, 1]$

Find the characteristic function of X.

Find $\overline{X}$ and σ_X^2

Note: See problems 3.14 and 3.15 in Proakis and Salehi

Exercise:

Problem:

Consider outcomes of a fair dice $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$. Define events $A = \{\omega, \omega | \text{ an even number appears}\}$ and $B = \{\omega, \omega | \text{ a number less than 5 appears}\}$. Are these events disjoint? Are they independent? (Show your work!)

Exercise:

Problem: This is problem 3.5 in Proakis and Salehi.

An information source produces 0 and 1 with probabilities 0.3 and 0.7, respectively. The output of the source is transmitted via a channel that has a probability of error (turning a 1 into a 0 or a 0 into a 1)

equal to 0.2.

What is the probability that at the output a 1 is observed?

What is the probability that a 1 was the output of the source if at the output of the channel a 1 is observed?

Exercise:

Problem:

Suppose X and Y are each Gaussian random variables with means μ_X and μ_Y and variances σ_X^2 and σ_Y^2. Assume that they are also independent. Show that $Z = X + Y$ is also Gaussian. Find the mean and variance of Z.

Homework 2 of Elec 430

Elec 430 homework set 2. Rice University Department of Electrical and Computer Engineering.

Problem 1

Suppose A and B are two Gaussian random variables each zero mean with

$A^2 < \infty$ and $B^2 < \infty$. The correlation between them is denoted by AB. Define the random process $X_t = A + Bt$ and $Y_t = B + At$.

- a) Find the mean, autocorrelation, and crosscorrelation functions of X_t and Y_t.
- b) Find the 1st order density of X_t, $f_{X_t}(x)$
- c) Find the conditional density of X_{t_2} given X_{t_1}, $f_{X_{t_2}\ X_{t_1}}(x_2|x_1)$. Assume $t_2 > t_1$

Note: see Proakis and Salehi problem 3.28

- d) Is X_t wide sense stationary?

Problem 2

Show that if X_t is second-order stationary, then it is also first-order stationary.

Problem 3

Let a stochastic process X_t be defined by $X_t = \cos(\Omega t + \Theta)$ where Ω and Θ are statistically independent random variables. Θ is uniformaly distributed over $[-\pi, \pi]$ and Ω has an unknown density $f_\Omega(\omega)$.

- a) Compute the expected value of X_t.
- b) Find an expression for the correlation function of X_t.
- c) Is X_t wide sense stationary? Show your reasoning.
- d) Find the first-order density function $f_{X_t}(x)$.

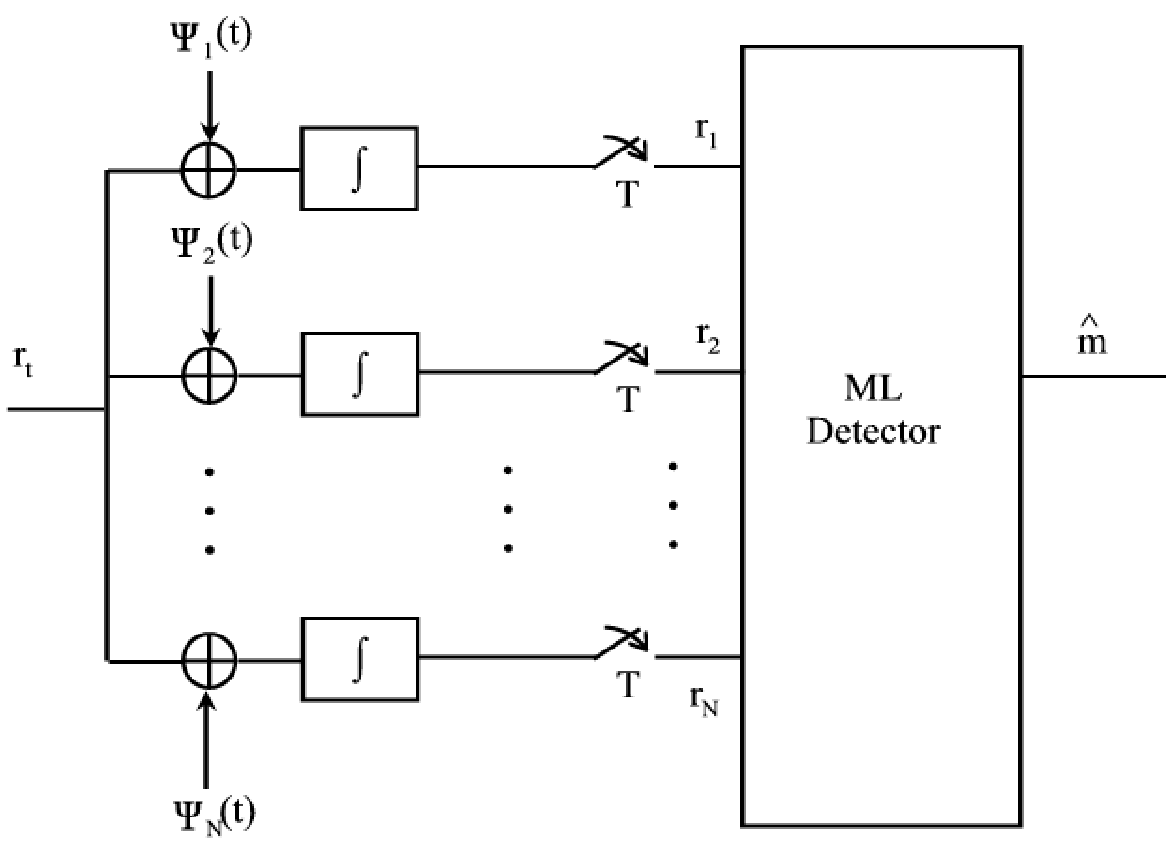
Homework 5 of Elec 430

Problem 1

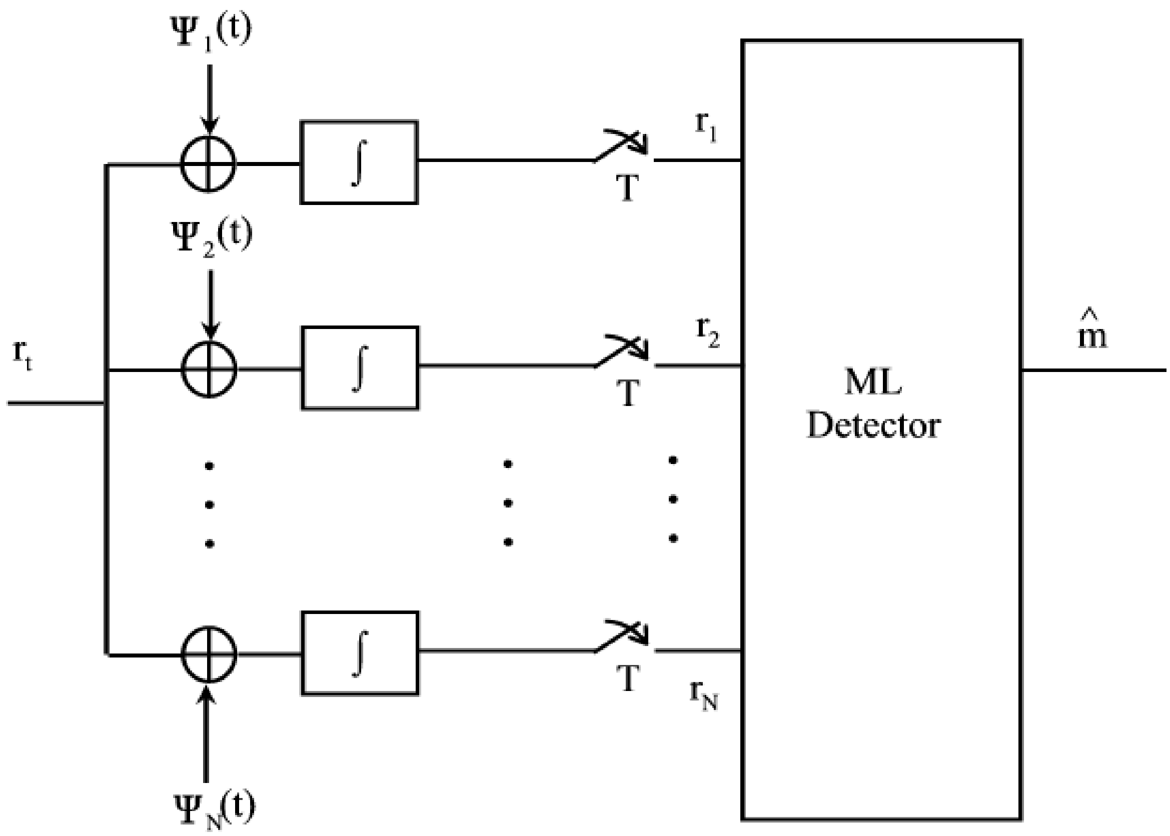
Consider a ternary communication system where the source produces three possible symbols: 0, 1, 2.

a) Assign three modulation signals $s_1(t)$, $s_2(t)$, and $s_3(t)$ defined on $t \in [0, T]$ to these symbols, 0, 1, and 2, respectively. Make sure that these signals are not orthogonal and assume that the symbols have an equal probability of being generated.

b) Consider an orthonormal basis $\psi_1(t)$, $\psi_2(t)$, ..., $\psi_{\mathrm{N}}(t)$ to represent these three signals. Obviously N could be either 1, 2, or 3.



Now consider two different receivers to decide which one of the symbols were transmitted when $r_t = s_m(t) + N_t$ is received where $m = \{1, 2, 3\}$ and N_t is a zero mean white Gaussian process with $S_N(f) = \frac{N_0}{2}$ for all f. What is $f_{\boldsymbol{r}|s_m(t)}$ and what is $f_{\boldsymbol{Y}|s_m(t)}$?



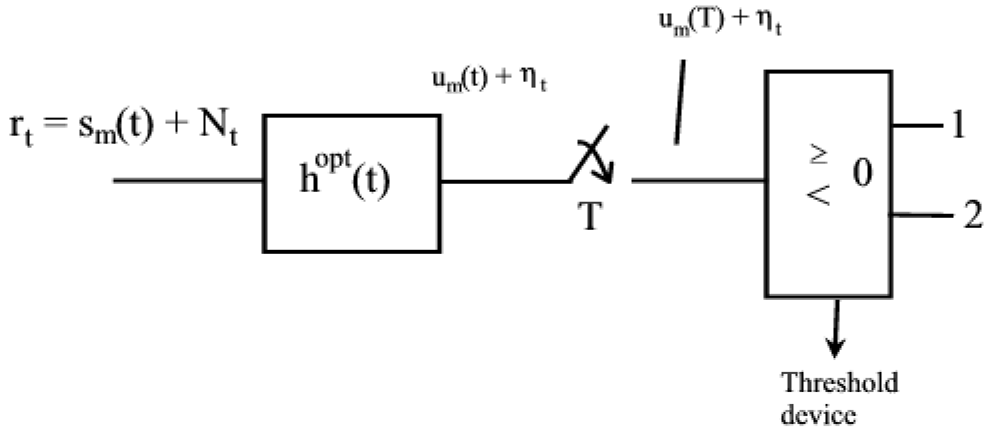
Find the probability that $\widehat{m} \neq m$ for both receivers. $P_e = \Pr\left[\widehat{m} \neq m\right]$.

Problem 2

Proakis and Salehi problems 7.18, 7.26, and 7.32

Problem 3

Suppose our modulation signals are $s_1(t)$ and $s_2(t)$ where $s_1(t) = e^{-t^2}$ for all t and $s_2(t) = -s_1(t)$. The channel noise is AWGN with zero mean and spectral height $\frac{N_0}{2}$. The signals are transmitted equally likely.

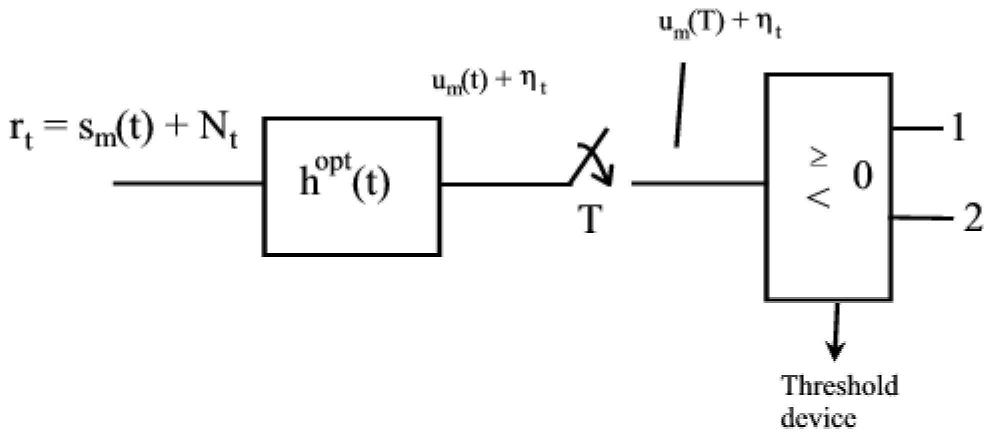


Find the impulse response of the optimum filter. Find the signal component of the output of the matched filter at $t = T$ where $s_1(t)$ is transmitted; i.e., $u_1(t)$. Find the probability of error $\Pr\left[\widehat{m} \neq m\right]$.

In this part, assume that the power spectral density of the noise is not flat and in fact is
Equation:

$$S_N(f) = \frac{1}{(2\pi f)^2 + \alpha^2}$$

for all f, where α is real and positive. Can you show that the optimum filter in this case is a cascade of two filters, one to whiten the noise and one to match to the signal at the output of the whitening filter?



c) Find an expression for the probability of error.

Homework 3 of Elec 430

Exercise:

Problem:

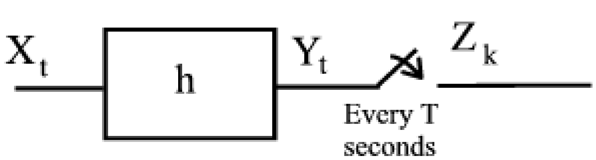
Suppose that a white Gaussian noise X_t is input to a linear system with transfer function given by

Equation:

$$H(f) = \begin{cases} 1 & \text{if } |f| \le 2 \\ 0 & \text{if } |f| > 2 \end{cases}$$

Suppose further that the input process is zero mean and has spectral height $\frac{N_0}{2} = 5$. Let Y_t denote the resulting output process.

1. Find the power spectral density of Y_t. Find the autocorrelation of Y_t (i.e., $R_Y(\tau)$).
2. Form a discrete-time process (that is a sequence of random variables) by sampling Y_t at time instants T seconds apart. Find a value for T such that these samples are uncorrelated. Are these samples also independent?
3. What is the variance of each sample of the output process?

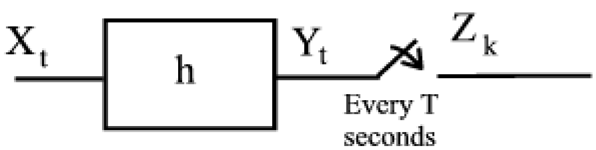


$Z_k = Y_{kT}$ for k = ...-1, 0, 1, 2, ...

Exercise:

Problem:

Suppose that X_t is a zero mean white Gaussian process with spectral height $\frac{N_0}{2} = 5$. Denote Y_t as the output of an integrator when the input is Y_t.



$Z_k = Y_{kT}$ for $k = \ldots -1, 0, 1, 2, \ldots$

1. Find the mean function of Y_t. Find the autocorrelation function of Y_t, $R_Y(t+\tau, t)$
2. Let Z_k be a sequence of random variables that have been obtained by sampling Y_t at every T seconds and dumping the samples, that is
   Equation:

   $$Z_k = \int_{(k-1)T}^{kT} X_\tau \, \mathrm{d}\, \tau$$

   Find the autocorrelation of the discrete-time processes Z_k's, that is, $R_Z(k+m, k) = E(Z_{\mathrm{k+m}} Z_k)$
3. Is Z_k a wide sense stationary process?

Exercise:

Problem: *Proakis and Salehi*, problem 3.63, parts 1, 3, and 4

Exercise:

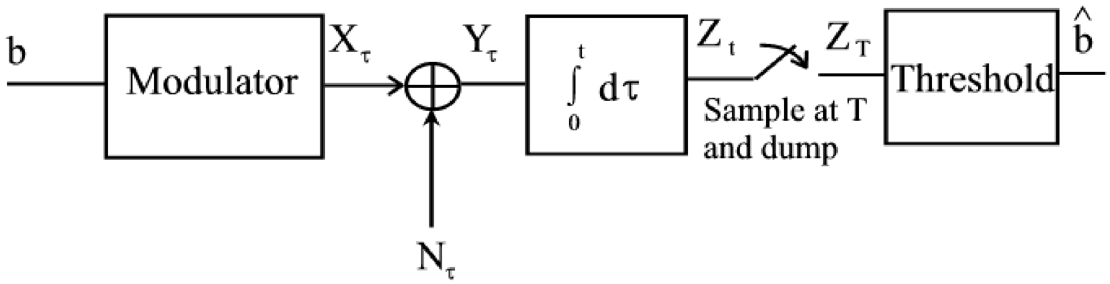
Problem: *Proakis and Salehi*, problem 3.54

Exercise:

Problem: *Proakis and Salehi*, problem 3.62

Exercises on Systems and Density
Exercise:

Problem: Consider the following system



Assume that N_τ is a white Gaussian process with zero mean and spectral height $\frac{N_0}{2}$.

If b is "0" then $X_\tau = Ap_T(\tau)$ and if b is "1" then $X_\tau = (-A)p_T(\tau)$ where $p_T(\tau) = \begin{cases} 1 \text{ if } 0 \leq \tau \leq T \\ 0 \text{ otherwise} \end{cases}$. Suppose $\Pr[b = 1] = \Pr[b = 0] = 1/2$.

1. Find the probability density function Z_T when bit "0" is transmitted and also when bit "1" is transmitted. Refer to these two densities as $\mathrm{f}_{Z_T,H_0}(z)$ and $\mathrm{f}_{Z_T,H_1}(z)$, where H_0 denotes the hypothesis that bit "0" is transmitted and H_1 denotes the hypothesis that bit "1" is transmitted.
2. Consider the ratio of the above two densities; i.e.,
   Equation:

$$\Lambda(z) = \frac{\mathrm{f}_{Z_T,H_0}(z)}{\mathrm{f}_{Z_T,H_1}(z)}$$

and its natural log $\ln(\Lambda(z))$. A reasonable scheme to decide which bit was actually transmitted is to compare $\ln(\Lambda(z))$ to a fixed threshold γ. ($\Lambda(z)$ is often referred to as the likelihood

function and $\ln(\Lambda(z))$ as the log likelihood function). Given threshold γ is used to decide $\hat{b} = 0$ when $\ln(\Lambda(z)) \geq \gamma$ then find $\Pr\left[\hat{b} \neq b\right]$ (note that we will say $\hat{b} = 1$ when $\ln(\Lambda(z)) < \gamma$).

3. Find a γ that minimizes $\Pr\left[\hat{b} \neq b\right]$.

Exercise:

Problem: *Proakis and Salehi*, problems 7.7, 7.17, and 7.19

Exercise:

Problem: *Proakis and Salehi*, problem 7.20, 7.28, and 7.23

Homework 6 of Elec 430

Homework set 6 of ELEC 430, Rice University, Department of Electrical and Computer Engineering

Problem 1

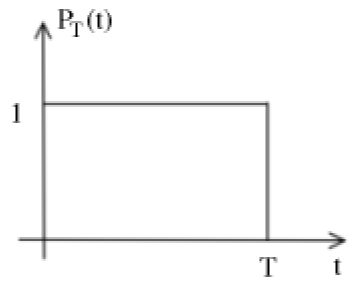
Consider the following modulation system
Equation:

$$s_0(t) = AP_T(t) - 1$$

and
Equation:

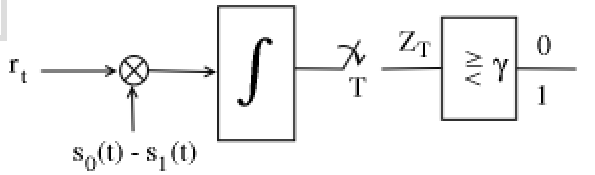
$$s_1(t) = (-(AP_T(t))) - 1$$

for $0 \leq t \leq T$ where $P_T(t) = \begin{cases} 1 & \text{if } 0 \leq t \leq T \\ 0 & \text{otherwise} \end{cases}$



The channel is ideal with Gaussian noise which is $\mu_N(t) = 1$ for all t, wide sense stationary with $R_N(\tau) = b^2 e^{-|\tau|}$ for all $\tau \in \mathbb{R}$. Consider the following receiver structure

$$\boxed{r_t = s_m(t) + N_t}$$



- a) Find the optimum value of the threshold for the system (e.g., γ that minimizes the P_e). Assume that $\pi_0 = \pi_1$
- b) Find the error probability when this threshold is used.

Problem 2

Consider a PAM system where symbols a_1, a_2, a_3, a_4 are transmitted where $a_{\text{n}} \in \{2A, A, -A, -(2A)\}$. The transmitted signal is

Equation:

$$X_t = \sum_{n=1}^{4} a_n s(t - nT)$$

where $s(t)$ is a rectangular pulse of duration T and height of 1. Assume that we have a channel with impulse response $g(t)$ which is a rectangular pulse of duration T and height 1, with white Gaussian noise with $S_N(f) = \frac{N_0}{2}$ for all f.

- a) Draw a typical sample path (realization) of X_t and of the received signal r_t (do not forget to add a bit of noise!)
- b) Assume that the receiver knows $g(t)$. Design a matched filter for this transmission system.
- c) Draw a typical sample path of Y_t, the output of the matched filter (do not forget to add a bit of noise!)
- d) Find an expression (or draw) $u(nT)$ where $u(t) = s{*}g{*}h^{\text{opt}}(t)$.

Problem 3

Proakis and Salehi, problem 7.35

Problem 4

Proakis and Salehi, problem 7.39

Homework 7 of Elec 430

Exercise:

Problem:

Consider an On-Off Keying system where $s \quad t \qquad A \qquad \pi f_c t \quad \theta$ for $\quad t \quad T$ and $s \quad t$ for $\quad t \quad T$. The channel is ideal AWGN with zero mean and spectral height $\frac{N}{\ }$.

1. Assume θ is known at the receiver. What is the average probability of bit-error using an optimum receiver?
2. Assume that we estimate the receiver phase to be θ and that $\theta \quad \theta$. Analyze the performance of the matched filter with the wrong phase, that is, examine P_e as a function of the phase error.
3. When does noncoherent become preferable? (You can find an expression for the P_e of noncoherent receivers for OOK in your textbook.) That is, how big should the phase error be before you would switch to noncoherent?

Exercise:

Problem: *Proakis and Salehi*, Problems 9.4 and 9.14

Exercise:

Problem:

A **coherent** phase-shift keyed system operating over an AWGN channel with two sided power spectral density $\frac{N}{\ }$ uses $s \quad t \qquad A p_T \ t \qquad \omega_c t \quad \theta$ and $s \quad t \qquad A p_T \ t \qquad \omega_c t \quad \theta$ where $\ i \ i \qquad \theta_i \qquad \frac{\pi}{\ }$, are constants and that $f_c T$ with $\omega_c \quad \pi f_c$.

1. Suppose θ and θ are **known** constants and that the optimum receiver uses filters matched to $s \quad t$ and $s \quad t$. What are the values of P_e and P_e ?

2. Suppose θ and θ are **unknown** constants and that the receiver filters are matched to $s \quad t \qquad Ap_T \; t \qquad \omega_c t$ and $s \quad t \qquad Ap_T \; t \qquad \omega_c t \quad \pi$ and the threshold is zero.

Note: Use a correlation receiver structure.

What are P_e and P_e now? What are the minimum values of P_e and P_e (as a function of θ and θ)?

Homework 8 of Elec 430

Exercise:

Problem: *Proakis and Salehi*, Problems 9.15 and 9.16

Exercise:

Problem: *Proakis and Salehi*, Problem 9.21

Exercise:

Problem: *Proakis and Salehi*, Problems 4.1, 4.2, and 4.3

Exercise:

Problem: *Proakis and Salehi*, Problems 4.5 and 4.6

Homework 9 of Elec 430

Exercise:

Problem: *Proakis and Salehi*, Problems 4.22 and 4.28

Exercise:

Problem: *Proakis and Salehi*, Problems 4.21 and 4.25

Exercise:

Problem: *Proakis and Salehi*, Problems 10.1 and 10.6

Exercise:

Problem: *Proakis and Salehi*, Problems 10.8 and 10.9

Exercise:

Problem:

For this problem of the homework, please either make up a problem relevant to chapters 6 or 7 of the notes or find one from your text book or other books on Digital Communication, state the problem clearly and carefully and then solve.

> **Note:** If you would like to choose one from your textbook, please reserve your problem on the white board in my office. (You may not pick a problem that has already been reserved.)

Please write the problem and its solution on separate pieces of paper so that I can easily reproduce and distribute them to others in the class.